The Effect of Perceptual Salience on Phonetic Accommodation in Cross-Dialectal Conversation in Spanish

by

Bethany MacLeod

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Department of Linguistics
University of Toronto

© Copyright by Bethany MacLeod 2012
The Effect of Perceptual Salience on Phonetic Accommodation in Cross-Dialectal Conversation in Spanish

Bethany MacLeod

Doctor of Philosophy

Department of Linguistics
University of Toronto

2012

Abstract

Phonetic accommodation is the process whereby speakers in an interaction modify their speech in response to their interlocutor. The social-psychological theory of Communication Accommodation Theory (Giles 1973) predicts that speakers will converge towards (become more similar to) their interlocutors in order to decrease social distance, whereas they will diverge from (become less similar to) their interlocutors to accentuate distinctiveness or show disdain. Previous studies have found that phonetic accommodation is affected by many social, situational and linguistic factors (Abrego-Collier et al. 2011; Black 2012; Babel 2009, 2010, 2012; Babel et al. 2012; Kim, Horton & Bradlow 2011; Nielsen 2011; Pardo et al. 2012). With respect to accommodation across dialects, a handful of studies have suggested that the perceptual salience of the various differences between two dialects might affect the pattern; however, these studies make conflicting predictions. Trudgill (1986) predicts that speakers will converge more towards the more salient dialectal differences, while Kim et al. (2011) and Babel (2009, 2010) suggest the opposite: that speakers will converge on the less salient differences.

This thesis investigates how the perceptual salience of 6 differences between Buenos Aires Spanish and Madrid Spanish affect the pattern of phonetic accommodation in conversation. The results are considered both in terms of the magnitude of the changes that the participants make as well as the direction of the change (convergence or divergence). The
results show that perceptual salience has a significant effect on the magnitude of the change, with all participants making greater changes as perceptual salience increases. On the other hand, perceptual salience was found not to have a consistent effect for all speakers on the likelihood of converging or diverging on the dialectal differences. I argue that the lack of consistent effect of salience on the direction of the change stems from individual differences in motivation to take on the opposing dialect norms and issues of personal identity, whereas the very consistent effect of salience on the magnitude of the change suggests that there is something more basic or systematic about how salience interacts with the extent to which speakers accommodate.
Acknowledgments

Throughout my entire career as a graduate student I’ve been very lucky; I’ve had lots of good things happen to me and very few bad things. One of the best things was the string of excellent professors I’ve had teaching me and advising me. In particular, I would like to thank Yoonjung Kang for being my advisor. She is open-minded, flexible, dependable, generous, encouraging, and incredibly smart. She has always given me the help and guidance I needed in whatever way I needed it, but also allowed me to explore my own ways of doing things. I hope I can be as good an advisor as Yoonjung when I grow up.

The rest of my committee and other professors have also helped me immeasurably. I would like to thank Alexei Kochetov for his support in being involved in all of my major projects during my PhD as well as giving me the chance to be involved with the lab and to know what it’s like to wear a false palate. I would like to thank Laura Colantoni for being on my advisory committees for every single major project I’ve done in grad school, including during my MA. I hope she can see some improvement in me since my forum paper. I would also like to thank Keren Rice for her enduring encouragement in all areas of academics and life in general. Furthermore, I’d like to thank Laura and Keren for encouraging me to go abroad to improve my Spanish, since without having done that I would not have managed to secure the job I have, I would not have started TorontoBabel, and I would have far fewer stories of embarrassing myself in Spanish. I’d also like to thank Jack Chambers and José Ignacio Hualde for serving on my committee at my defense. Their insights and discussion were interesting and also fun. Thanks also to Molly Babel for her help at various stages of this project and for her support, in general. I thank Mary Hsu and Jill Given-King for their friendship and for all their help in keeping things running smoothly on the admin side. Elaine Gold deserves a huge thank you for half-adopting me and allowing me to live part-time at her home. Thank you also to Chandan Narayan for all his encouragement and friendship.

Another way I’ve been very lucky in grad school is to be surrounded by a diverse, interesting, hilarious group of fellow students. I’d like to especially thank Ailis Cournane, Sarah Clarke, and Liisa Duncan for all their time and energy spent chatting, laughing, encouraging, plotting, gossiping, justifying, listening, commiserating, imagining, congratulating, and sometimes drinking. In addition, I’d like to thank Yasaman Rafat, a graduate of the
Department of Spanish and Portuguese, for her support and insight over the years. As well, I’d like to thank Stephanie Dixon and Alison Luby for their encouragement and understanding as my non-linguistic academic friends. Outside of the university, I am very thankful to all the amazing friends I’ve made through TorontoBabel, especially Eduardo Costa, for all the laughs that were an excellent weekly break from my thesis.

The most fundamental way I’ve been lucky is in having an incredible family. My parents deserve an enormous thank you. I definitely could not have gotten through my PhD without their support, of both the financial and emotional kind, so thank you very much! And last, but certainly not least, a great big thank you to Conrad for always being there, being able to give me encouragement and confidence without patronizing, for managing to keep himself occupied while I spent all my time writing, and for bringing me a cider right when I needed one.

To everybody, thank you.
# Table of Contents

ACKNOWLEDGMENTS IV  
TABLE OF CONTENTS VI  
LIST OF TABLES IX  
LIST OF FIGURES X  
LIST OF APPENDICES XII  

CHAPTER 1 INTRODUCTION 1  

CHAPTER 2 LONG-TERM AND SHORT-TERM DIALECT ACCOMMODATION 7  
2.1 Long-Term Dialect Accommodation 7  
2.1.1 Change-by-Accommodation Model 8  
2.1.2 Second-Dialect Acquisition 9  
2.1.2.1 Second-Dialect Acquisition in Adults 10  
2.1.2.2 Second-Dialect Acquisition in Children 13  
2.2 Short-Term Phonetic Accommodation 15  
2.2.1 Communication Accommodation Theory 15  
2.2.2 Phonetic Imitation in Non-Social Contexts 17  
2.2.3 Factors Affecting Phonetic Imitation 19  
2.2.3.1 Social and Situational Factors 19  
2.2.3.2 Linguistic Factors 22  

CHAPTER 3 PERCEPTUAL SALIENCE 26  
3.1 Salience as a List of Criteria 29  
3.1.1 Salience in Short- and Long-Term Accommodation 29  
3.1.2 Objective and Subjective Criteria for Salience 34  
3.1.3 Extra-Linguistic Factors for Salience 37  
3.2 Evaluation of Criteria-List Definition of Salience 39  
3.2.1 Social Context 40  
3.2.2 Individual Variation in Perception 44  
3.2.3 Salience as a Gradient Concept 46  
3.3 Salience as a Measurable Property of Linguistic Variables 47  
3.4 Salience and Phonetic Accommodation 50  

CHAPTER 4 THE DIALECTS: BUENOS AIRES SPANISH AND MADRID SPANISH 53  
4.1 Dialectal Difference #1 (BAS /s/ versus MS /θ/) 57  
4.2 Dialectal Difference #2 (BAS /ʃ/ versus MS /ʝ/) 58  
4.3 Dialectal Difference #3 (BAS laminal [s] versus MS apical [ʃ]) 60  
4.4 Dialectal Difference #4 (BAS [x] versus MS [χ]) 62  
4.5 A Note on Spanish Vocalic Sequences 63  
4.6 Dialectal Difference #5 (Diphthongization of Mid Vowel Sequences in BAS, but not in MS) 64  
4.7 Dialectal Difference #6 (Potential Exceptional Hiatus in MS, but not in BAS) 65
CHAPTER 5 METHODOLOGY

5.1 Participants

5.2 Outline of Tasks

5.2.1 Step 1: Baseline (Pre-Conversation) Production Task

5.2.1.1 Stimuli: Fricatives

5.2.1.2 Stimuli: Vocalic Sequences

5.2.2 Step 2: Perception Task

5.2.2.1 Perception Task Stimuli

5.2.3 Step 3: Repetition Task

5.2.4 Step 4: Conversation

5.2.4.1 Map Task

5.2.4.2 Guided Conversation

5.2.5 Step 5: Post-Conversation Word List

5.2.6 Step 6: Questionnaire

5.3 Data Analysis

5.3.1 Acoustic Analysis

5.3.1.1 Relative Intensity

5.3.1.2 Centre of Gravity

5.3.1.3 Mean Onset F2

5.3.1.4 Vocalic Sequence Duration

5.3.1.5 Composite Sequence Duration + Vocoid Height

5.3.2 Assessing Phonetic Accommodation

5.3.2.1 A Note on Within-Category and Across-Category Accommodation

5.3.3 Standardizing the Magnitude of the Change

5.3.4 Mixed-Effects Modeling in R

CHAPTER 6 RESULTS

6.1 Production of Dialectal Differences Pre-Conversation

6.1.1 Dialectal Difference #1: BAS /s/ versus MS /θ/

6.1.2 Dialectal Difference #2: BAS /ʃ/ versus MS /ʝ/

6.1.3 Dialectal Difference #3: BAS laminal [s] versus MS apical [s̺]

6.1.4 Dialectal Difference #4: BAS Velar [x] versus MS Uvular [χ]

6.1.5 Dialectal Difference #5: Diphthongization of Mid Vowel Sequences in BAS, but not in MS

6.1.6 Dialectal Difference #6: Exceptional Hiatus in MS, but not in BAS

6.1.7 Summary of Pre-Conversation Production of Dialectal Differences

6.1.8 Repetition of Opposing Dialect Variants

6.2 Perceptual Salience of the Dialectal Differences

6.2.1 Control Trials

6.2.2 Perceptual Salience of the Dialectal Differences

6.2.3 The Unexpected Salience of Difference #3

6.2.4 Interspeaker Variation

6.3 The Pattern of Phonetic Accommodation

6.3.1 The Magnitude of Change

6.3.1.1 Perceptual Salience and the Magnitude of Change

6.3.2 The Direction of Change

6.3.2.1 Perceptual Salience and the Direction of Change

6.4 Summary of Results
List of Tables

Table 1: Trudgill (1986)’s four criteria for salience of a linguistic variable 30
Table 2: Objective and subjective criteria for salience (Auer et al. 1998) 35
Table 3: Consonantal phonemic inventory of Buenos Aires Spanish 54
Table 4: Consonantal phonemic inventory of Madrid Spanish 55
Table 5: List of dialectal differences between Madrid Spanish and Buenos Aires Spanish 57
Table 6: Gender and ages of both participants per pair, and length of time in Madrid for the BAS speakers. 68
Table 7: Breakdown of stimuli included in experiment for Differences #1 - #4 70
Table 8: Breakdown of stimuli included in experiment for Differences #5 - #6 71
Table 9: Breakdown of stimuli included in perception task 76
Table 10: Control stimuli used in perceptual task to determine perceptual salience. 77
Table 11: Standardizing the MS speaker in Pair 1’s changes made on Difference #1 88
Table 12: Standardizing the MS speaker in Pair 1’s changes made on Difference #3 88
Table 13: Mean %-Correct by dialectal difference and dialect, and overall mean 117
Table 14: Model summary: logistic mixed effects model with Accuracy as dependant 118
Table 15: Results of post-hoc tests for salience of the dialectal differences for BAS speakers 119
Table 16: Results of post-hoc tests for salience of the dialectal differences for MS speakers 120
Table 17: Range, mean, and median accuracy rates by dialectal difference for BAS speakers. 126
Table 18: Range, mean, and median accuracy rates by dialectal difference for MS speakers. 126
Table 19: Model summary: linear mixed effects model with LOG ABSCHANGE as dependant 144
Table 20: By-subject coefficients for %CORRECT 145
Table 21: Model summary: linear mixed effects model with DIRECTION as dependant 153
Table 22: By-subject coefficients for %CORRECT 154
Table 23: Stimuli for Dialectal Difference #1: orthographic <z> 204
Table 24: Stimuli for Dialectal Difference #1: orthographic <c> 205
Table 25: Stimuli for Dialectal Difference #2: orthographic <y> 206
Table 26: Stimuli for Dialectal Difference #2: orthographic <ll> 207
Table 27: Stimuli for Dialectal Difference #3 208
Table 28: Stimuli for Dialectal Difference #4 209
Table 29: Stimuli for Dialectal Difference #5 210
Table 30: Stimuli for Dialectal Difference #6 211
Table 31: Model summary: linear mixed effects model with relative intensity as dependant. 220
Table 32: Model summary: linear mixed effects model with relative intensity as dependant. 220
Table 33: Model summary: linear mixed effects model with centre of gravity as dependant. 220
Table 34: Model summary: linear mixed effects model with F2 at the onset of the following vowel as dependant. 220
Table 35: Model summary: linear mixed effects model with composite measure (sequence duration and F1) as dependant. 221
Table 36: Model summary: linear mixed effects model with sequence duration as dependant. 221
List of Figures

Figure 1: The phonemic vowel system of Spanish 54
Figure 2: Outline of tasks performed by each participant in the experiment 68
Figure 3: Sample trial of the perception task to determine perceptual salience 73
Figure 4: Difference #1 - Distribution of relative intensity in dB of BAS /s/ and MS /θ/ 92
Figure 5: Difference #1 - Mean relative intensity in dB of BAS /s/ and MS /θ/ by individual participant 93
Figure 6: Difference #2 - Distribution of relative intensity in dB of BAS /ʃ/ and MS /ʝ/ 94
Figure 7: Difference #2 - Mean relative intensity in dB of BAS /ʃ/ and MS /ʝ/ by individual participant 95
Figure 8: Difference #3 - Distribution of centre of gravity of /s/ in Hertz by dialect 96
Figure 9: Mean centre of gravity in Hertz of /s/ by individual participant 96
Figure 10: Difference #4 - Distribution of F2 in mels at onset of vowel following /ʃ/ by dialect 98
Figure 11: Mean F2 in mels at onset of vowel following /ʃ/ by individual participant 99
Figure 12: Difference #5 - Distribution of normalized duration of vowel sequence in milliseconds by sequence type and by dialect 101
Figure 13: Mean normalized duration of mid vowel sequence in milliseconds by individual participant 102
Figure 14: Difference #5 - Mean F1 in Mels of first vowel in a vowel sequence by sequence type and by dialect 103
Figure 15: Mean F1 in Mels of the first vocoid in an /eV/ sequence by individual participant 104
Figure 16: Difference #5 - Mean F2 in Mels of first vocoid in a vowel sequence by sequence type and by dialect 105
Figure 17: Equation to predict dialect of a given vocalic sequence based on duration of the sequence and F1 of the first vocoid 106
Figure 18: Difference #6 - Mean normalized duration of vocalic sequences that are typically syllabified as a diphthong versus those that could be syllabified as an exceptional hiatus by dialect 109
Figure 19: Mean normalized duration of potential exceptional hiatus in milliseconds by individual participant 110
Figure 20: Percentage of responses to the control trials by participants’ dialect and accuracy 114
Figure 21: Percentage of correct responses to the control trials by individuals 115
Figure 22: Percentage of correct responses by dialectal difference by dialect of the participants 116
Figure 23: Illustration of the perceptual salience of the 6 dialectal differences for the BAS speakers 119
Figure 24: Illustration of the perceptual salience of the 6 dialectal differences for the MS speakers 120
Figure 25: Mean centre of gravity in Hertz for pre-conversation MS and BAS /s/ by following vowel 123
Figure 26: Distribution of change from pre- to post-conversation by BAS speakers by dialectal difference 128
Figure 27: Distribution of change from pre- to post-conversation by MS speakers by dialectal difference 129
Figure 28: Plot of change (magnitude + direction) by perceptual salience including all speakers 130
Figure 29: Amount of overlap in pre-conversation (left panel) and post-conversation (right panel) between BAS speaker in Pair 6 and entire MS group for Difference #1 132
Figure 30: Amount of overlap in pre-conversation (left panel) and post-conversation (right panel) between BAS speaker in Pair 4 and entire MS group for Difference #3 133
Figure 31: Distribution of AbsChange (the absolute value of the change from pre- to post-conversation) across all participants and all dialectal differences 134
Figure 32: Quantile-quantile plot of AbsChange (the absolute value of the change from pre- to post-conversation) across all participants and all dialectal differences 135
Figure 33: Distribution of the log-transformed AbsChange (the absolute value of the change from pre- to post-conversation) across all participants and all dialectal differences 136
Figure 34: Quantile-quantile plot of log-transformed AbsChange (the absolute value of the change from pre- to post-conversation) across all participants and all dialectal differences 136
Figure 35: Mean standardized absolute change on all dialectal differences (excluding Difference #3) by dialect 137
Figure 36: Mean standardized absolute change from pre- to post- conversation by gender and dialect 138
Figure 37: Scatterplot of the log-transformed absolute value of the change and log lexical frequency for all participants, all dialectal differences (excluding Difference #3), and all tokens 139
Figure 38: Scatterplot of the log-transformed absolute value of the change and %-overlap for all participants, all dialectal differences (excluding Difference #3), and all tokens

Figure 39: Mean standardized absolute change from pre- to post- conversation by previous exposure to token (N = not included in map task, Y= included in map task) by dialect

Figure 40: Mean standardized absolute change from pre- to post- conversation on tokens converged upon and diverged upon by dialect

Figure 41: Scatterplot of the log-transformed absolute value of the change perceptual salience (%-Correct) for all participants and all dialectal differences (excluding Difference #3) by dialect

Figure 42: Proportion of tokens converged upon by dialect

Figure 43: Proportion of tokens converged upon by dialect by gender

Figure 44: Proportion of tokens converged upon by dialectal difference by dialect

Figure 45: Scatterplot of the likelihood of converging (DIV=0, CON=1) by the log-transformed lexical frequency for all participants and all dialectal differences (excluding Difference #3)

Figure 46: Scatterplot of the likelihood of converging (DIV=0, CON=1) by %-overlap for all participants and all dialectal differences (excluding Difference #3)

Figure 47: Proportion of tokens converged upon by dialect by presence of token in the map task (N = not present, Y = present)

Figure 48: Scatterplot of the likelihood of converging (DIV=0, CON=1) by perceptual salience (%-Correct) for all participants and all dialectal differences (excluding Difference #3) by dialect

Figure 49: Scatterplot of the log-transformed absolute value of the change from pre- to post-conversation by %-overlap for all participants and all dialectal differences (excluding Difference #3)

Figure 50: Scatterplot of the log-transformed absolute value of the change from pre- to post-conversation by %-overlap for all participants and all dialectal differences for all tokens converged upon (excluding Difference #3)

Figure 51: Proportion of trials converged upon by gender and dialectal difference

Figure 52: Proportion of trials correctly responded to in the perception task by gender and dialectal difference

Figure 53: The standardized magnitude of the change from pre- to post-conversation on Difference #3 by dialect

Figure 54: The standardized magnitude of the change from pre- to post-conversation by dialectal difference by dialect

Figure 55: The proportion of tokens converged upon on Difference #3 by dialect

Figure 56: Boxplot of centre of gravity of BAS /ʃ/, MS /s/, and BAS /s/
List of Appendices

Appendix A – Stimuli
Dialectal Difference #1 (BAS /s/ vs. MS /θ/): orthographic <z> 204
Dialectal Difference #1 (BAS /s/ vs. MS /θ/): orthographic <c> 205
Dialectal Difference #2 (BAS /ʃ/ vs. MS /ʝ/): orthographic <y> 206
Dialectal Difference #2 (BAS /ʃ/ vs. MS /ʝ/): orthographic <ll> 207
Dialectal Difference #3 (BAS Laminal [s] vs. MS Apical [ʒ]): orthographic <s> 208
Dialectal Difference #4 (BAS Velar [ʃ] vs. MS Uvular [χ]) 209
Dialectal Difference #5 (Diphthongization of Mid Vowel Sequences in BAS, but not in MS) 210
Dialectal Difference #6 (Potential Exceptional Hiatus in MS, but not in BAS) 211

Appendix B – Information Letter and Consent Form
Spanish 212
English 214

Appendix C – Background Questionnaire
Spanish – Buenos Aires Spanish speakers 216
Spanish – Madrid Spanish speakers 217
English – Buenos Aires Spanish speakers 218
English – Madrid Spanish speakers 219

Appendix D – Statistical Analysis of Perception/Repetition Task Stimuli 220

Appendix E – Raw Data for Pre and Post-Conversation Production and Perception 222
Chapter 1
Introduction

On March 28, 2012, in an episode of the satirical news show *The Daily Show with Jon Stewart*, the host engaged in conversation with his guest, Pakistani journalist Ahmed Rashid. The topic of conversation was essentially the Taliban in Afghanistan and Pakistan. In North American English, the most typical vowel phoneme used to realize the vowel in the final syllable of the words *Taliban, Pakistan* and *Afghanistan* is the front /æ/ (realized as nasal because of the following nasal consonant). In the speech of Ahmed Rashid, however, these vowels were realized as closer to the back vowel /ɑ/, due to his being a speaker of a different dialect of English. After three minutes of conversation, Mr. Rashid gave a 16-second answer to a question of Jon Stewart’s, in which he used the words *Taliban, Pakistan* or *Afghanistan* (where the final syllable was realized with a back vowel) a total of 8 times, or once every 2 seconds, on average. When Jon Stewart began to respond, he stumbled slightly as he began to say *Afghan Taliban* and then proceeded to pronounce both final vowels as back vowels as Mr. Rashid does, rather than his dialect norm of the front vowel. Jon Stewart used a back vowel where he normally would not, in response to having perceived his interlocutor using the vowel in that context. This is an example of phonetic accommodation or convergence in cross-dialectal conversation, and it exemplifies the main topic of this thesis.

The study described in this thesis aims to determine how the perceptual salience of six phonetic or phonological differences between two dialects of Spanish (Buenos Aires Spanish and Madrid Spanish) affects the pattern of phonetic accommodation that takes place as a result of two speakers (one from Buenos Aires and one from Madrid) engaging in conversation together. Phonetic accommodation (also known as convergence, alignment or imitation) is the adjustment of the acoustic-phonetic properties of speech in response to exposure to another speaker. Accommodation or imitation has been observed at many levels of human behaviour, going far beyond merely imitation of the phonetic characteristics of speech. This tendency to imitate behaviours is known as ‘the chameleon effect’ and has been observed in movements such as smiling, face-rubbing, foot-shaking, and leg-crossing (Chartrand & Bargh 1999), and has been shown to begin in infancy, even with newborns (Meltzoff & Moore 1983). In addition, various aspects of speech have been found to be subject to alignment such as lexical choices (Garrod & Doherty 1994), grammatical structure (Bock 1986; Pickering and Ferreira 2008),
pitch and intensity levels (Goldinger 1998), speech rate (Giles, Coupland & Coupland 1991; Bosshardt, Sappok, Knipschild & Hölscher 1997), and acoustic characteristics of sounds, such as vowel spaces (Delvaux & Soquet 2007; Evans & Iverson 2007; Babel 2012) and voice onset time (Shockley, Sabadini & Fowler 2004; Abrego-Collier, Grove, Sonderegger & Yu 2011; Nielsen 2011). These studies have shown that talkers can affect the speech of others in both measurable and perceptible ways.

Why is it that we imitate the behaviours of those around us? Focusing specifically on speech accommodation, this question has two different facets: does imitating the speech of another serve a purpose? And what is the mechanism by which imitation occurs? Concentrating on the second part of this question, the literature on imitation has proposed two very different accounts of the actual mechanism behind the process. The first, based on the social-psychological framework of Communication Accommodation Theory (CAT: Giles 1973) is that the main driving force behind accommodation is social in that speakers adjust their speech in response to an interlocutor in order to manage social relationships online. These responses have two possible directions: speakers converge towards their interlocutor, that is, they come to speak more similarly, or they diverge away from their interlocutor meaning that they come to speak less similarly. Speakers are predicted to converge towards an interlocutor in order to minimize social distance or to diverge away from an interlocutor to show disdain or accentuate distinctiveness (Shepard, Giles & Le Poire 2001). Relating this approach to the first part of the question posed above, studies have found that there are social consequences to converging or diverging. Chartrand & Bargh (1999) suggest that “individuals use behaviour mimicry as a communication tool on a completely nonconscious level” (Chartrand & Bargh 1999: 901). Specifically, the authors investigated the extent to which mirroring of behaviours contributes to “liking” of the mimicker and a general sense of smoothness and success of the interaction. They found that participants whose movements were mirrored by the other participant, who was a research assistant for the study, judged the conversation to have proceeded more smoothly and liked the other participant more than those in the control group. So it seems that accommodation, both linguistic and non-linguistic, has social implications for how an individual or a situation will be perceived.
These social implications do not necessarily imply that the driving force behind accommodation is social, however. The other main account of the mechanism behind phonetic accommodation posits that there is a direct link between the perceptual and production systems that gives rise inevitably to imitation. Under this account, speech is perceived in terms of articulatory gestures that are, in turn, used as parameters for speech production, a relation that results in automatic and unintentional imitation (Goldinger 1998; Pickering & Garrod 2004; Shockley et al. 2004). Evidence for the automatic account of imitation comes from studies of imitation in non-social situations, such as in a lab where participants are only exposed to pre-recorded voices, where there would be little social motivation to converge, and yet convergence towards the model talker’s voice still occurs (Goldinger 1998; Shockley et al. 2004; Nielsen 2011). As noted by Black (2012), although these two approaches to explaining accommodation are frequently discussed as though they were in opposition, they may actually not be mutually exclusive since they seem to explain different aspects of the process. Babel (2009) concludes that accommodation may be automatic in the sense that a close link between perception and production triggers the process, but that it is mediated by various social and linguistic factors, which will be discussed in §2.

The main purpose of this thesis is not to debate the social or automatic explanation of accommodation but rather to explore a particular potentially mediating factor that has thus far not been considered explicitly in relation to accommodation. This factor is the perceptual salience of the various differences between two dialects. In the example given in the first paragraph, Jon Stewart momentarily took on the pronunciation of his conversation partner who spoke a different dialect of English. Between any two dialects of a language there will be various differences. In this thesis, I focus on the differences in phonological inventory or articulation of shared segments, which I refer to as dialectal differences (or simply differences) throughout the paper. Such dialectal differences will likely vary in how perceptually salient they are for speakers of the two dialects. Salience, in its most basic sense, “refers to the characteristic of being easily noticeable, prominent or conspicuous” (Siegel 2010: 129). One dialectal difference may be more perceptually salient, or noticeable, than another. How might this variation in salience affect the pattern of phonetic accommodation that takes place as a result of speakers of two different dialects engaging in conversation together? Will speakers be more likely to converge towards the differences that are most salient? Or would such a change be construed as too great a departure from their usual way of speaking?
No study has explicitly considered this question before, but a handful has suggested that salience may play a role; however, the nature of that role is far from clear. Trudgill (1986: 11) predicts that the most salient dialectal differences will be the ones to exhibit the greatest convergence. Presumably a speaker would need to notice a difference (if not consciously, then at least on some level) in order to be able to converge on it. The more salient a difference is, the more likely a speaker is to notice it. In this way, Trudgill’s prediction seems intuitive. However, as mentioned above, if a difference is very salient, taking on the opposing dialect’s pronunciation may seem to be a large shift from the normal speech pattern, potentially meaning that very salient differences might show less change. This is the finding of Babel (2009, 2010) and Kim, Horton & Bradlow (2011). In those studies, it was the least salient dialectal differences that were imitated most by the experimental participants, and the most salient differences showed less adjustment. Evans, Alshangiti, Hazan, Baker & Cyrus (in preparation) found that the greatest accommodation occurred on the most salient differences, but that the accommodation was in the direction of divergence since the participants were more likely to diverge on these differences, rather than converge as predicted by Trudgill (1986). In the case of these recent experimental studies, the focus was not on examining the role of perceptual salience, and as such, no metric of quantifying salience was used. Perhaps the diametrically opposed results can be attributed at least partly to differences in how salience was operationalized.

In §3, I discuss two main approaches used in the literature to define salience: the criteria-list approach, and the experimental approach. One of the main focuses and contributions of this thesis involves exploring previous methods of defining salience and arguing for the notion as a measurable property of linguistic variables that must be measured to be used as an explanatory factor. To determine the role of perceptual salience in cross-dialectal phonetic accommodation, I conducted an experiment testing 11 pairs of Spanish speakers (in each pair, one from Buenos Aires and one from Madrid). All of the testing took place in Madrid, where the Buenos Aires speakers had been living for various lengths of time. The experiment included a perception task designed to quantify the perceptual salience of six particular differences between the two dialects, based on the relevance of those six dialectal differences in assisting participants in recognizing their own dialect. In this way, perceptual salience is defined within the context of dialect recognition. In §3, I argue that considering salience within a particular context is crucial to developing a metric that has real value as a tool in explaining and predicting patterns in
speech production and perception. The experiment also included a production component, using the map task (Anderson, Bader, Bard, Boyle, Doherty, Garrod, Isard, Kowtko, McAllister, Miller, Sotillo, & Thompson 1991), in which a pair of speakers (one from each of the two dialect areas) engages in conversation together, thereby providing exposure to the other’s dialect. A word-list reading task flanked the conversation component and included lexical items containing the six dialectal differences under investigation. Acoustic analysis of the pre- and post-conversation word lists allowed me to determine the extent to which speakers changed their pronunciation as a result of having been exposed to the opposing dialect. Statistical analyses considered how the perceptual salience measure related to the extent and direction of the change in pronunciation, in order to establish how perceptual salience might mediate the process of accommodation alongside other social, situational, and linguistic factors that have been found to be relevant (discussed in §2).

The main finding is that perceptual salience does affect the pattern of accommodation in terms of the magnitude of the changes made in that as perceptual salience of the dialectal differences increases, the magnitude of the changes made also increases, and this effect was consistent even at the individual speaker level. In contrast, perceptual salience was not found to have a consistent effect on the direction of change made (convergence or divergence). Mixed effects models are used to determine the contribution of a variety of factors in explaining the variation. In addition, I appeal to social motivations and the need to maintain a phonological contrast to help explain the variation between convergence and divergence among and within speakers.

The remainder of the thesis is organized as follows. §2 provides the necessary background information about phonetic accommodation, both in terms of the long-term and short-term patterns. §3 discusses the criteria-list approach to defining perceptual salience and gives a critical evaluation of the method, as well as examining the experimental approach. Ultimately, I argue that salience is a measurable property of linguistic variables that must be considered within a particular context. §4 explains the six dialectal differences between Buenos Aires Spanish and Madrid Spanish and presents the acoustic correlates of the differences. §5 provides the experimental methodology and discusses how the data were analyzed acoustically and statistically. §6 presents the results of the acoustic and statistical analyses, giving the main results of the study. §7 contains a discussion of the significance of the findings and makes some
predictions about what the findings might mean for longer-term patterns of accommodation. Finally, §8 summarizes the study, provides the conclusions, and gives some possibilities for future research building on the results of this thesis.
Chapter 2  
Long-Term and Short-Term Dialect Accommodation

This chapter aims to contextualize the present study in the linguistic literature and to explain the gap in our current knowledge of accommodation that this thesis hopes to fill. I will summarize the important research that has been conducted on phonetic accommodation as well as point out that we are still left with many interesting questions, particularly with respect to the process of accommodation in cross-dialectal conversation. Improving our understanding of the path of accommodation across two different dialects of the same language is important because such knowledge will have implications for making predictions about longer term patterns of accommodation such as the acquisition of a second dialect and community-level changes in situations of dialect contact, such as dialect levelling.

§2.1 discusses the findings of studies of the longer term accommodation process of second-dialect acquisition in order to show how short-term accommodation can be situated in a broader context. §2.2 then gives the background on previous studies of short-term phonetic accommodation, explaining the main findings and the social, situational, and linguistic factors that affect the process. Within this section, I will highlight studies of short-term phonetic accommodation across two dialects and to illustrate that other factors may play a role in cross-dialectal accommodation.

2.1 Long-Term Dialect Accommodation

While the focus of this dissertation is phonetic accommodation in the short term, a discussion of dialect change and second-dialect acquisition showing the results of long-term contact between different linguistic systems is relevant since these longer term patterns may be related to the pattern in the short term.
2.1.1 Change-by-Accommodation Model

According to Trudgill (1986) and Chambers (1992), permanent changes to an adult’s pronunciation towards the characteristics of a second dialect (second-dialect acquisition or long-term phonetic accommodation) result from an accumulation of short-term convergences in face-to-face interactions with speakers of the second dialect. As mentioned in §1, experimental studies have shown that speakers can affect the speech of interlocutors and can come to sound more similar to each other in measurable and perceptible ways, even within a single conversation. During interactions with speakers of a second dialect, a transplanted speaker’s speech changes to either lose some of the characteristics of his native dialect (D1) and/or to take on some of the characteristics of the second dialect (D2), although within a single interaction such changes may be very small. As these small-scale changes accumulate over time, the changes become greater, and also become part of the transplanted speaker’s habitual way of speaking: she acquires (at least to some extent) aspects of the second dialect (Trudgill 1986; Chambers 1992).

This transition from short-term temporary convergence to long-term permanent convergence in the individual is predicted to be a precursor to community-level change, such as dialect levelling or sound change, more generally, under the Change-by-Accommodation model of language change (Niedzielski & Giles 1996). This model consists of three stages: 1. short-term accommodation in individual interactions, 2. long-term dialect accommodation resulting in permanent changes in the speech of individual speakers, and 3. the spread of new speech habits throughout the community. The following section discusses research that has considered the second stage of the Change-by-Accommodation model. Although this thesis focuses only on accommodation in the short term (within one conversation), many of the factors that affect acquisition of a second dialect also play a role in the short-term pattern, and as such are important to understand. Conceiving of short-term accommodation as the first step to community-level changes under the Change-by-Accommodation model allows short-term accommodation to be contextualized more broadly and to be related to more permanent changes in the individual (such as acquisition of a second dialect), or larger scale changes in entire speech communities.
2.1.2 Second-Dialect Acquisition

This subsection focuses on research centered on the second stage of the Change-by-Accommodation model: second-dialect acquisition. Sociolinguists, phoneticians, and laboratory phonologists have considered the path of acquisition of a second dialect (SDA) of one’s native language, although the amount of work focusing on SDA is dwarfed by that of the acquisition of a second language (SLA). Studies on SLA have found that after a certain age, acquiring second language (L2) phonology becomes more difficult such that native speakers of the L2 typically are able to perceive foreign accent in the learner’s speech, suggesting the presence of the so-called critical period for language acquisition (Birdsong 1999; Long 2007). However, many studies have also uncovered examples of relative success of L2 language learners, where even if the ultimate attainment of the L2 phonetic categories is not exactly native-like, production of L2 sounds is different from production of first language (L1) sounds, showing some degree of learning (Flege 1980, 1987, 1991; Steele 2001; Colantoni & Steele 2006, 2007; Munro & Derwing 2008; Menke & Face 2010).

Under Flege’s Speech Learning Model (SLM: Flege 1995), the production and perception of sounds is predicted to remain adaptable throughout the lifespan, suggesting that adults’ pronunciation does have the potential to change. If adults can learn to produce second-language categories at least somewhat in a native-like way, then adults should be capable of acquiring the phonetic categories of a second dialect as well. Of course, the details surrounding the situation of acquiring a second dialect can be very different from those of acquiring a second language. A major reason why people learn a second language is to be able to communicate with speakers of that language. Typically two dialects of the same language are mutually intelligible (Siegel 2010: 1), meaning that communication in its most basic sense is likely not the primary goal of SDA. Issues of comprehensibility can arise in cross-dialectal communication that stem from differences in lexical items and meanings, syntactic structures, phonological differences that cause homophony, etc., but these difficulties are usually relatively minor compared with potential problems in cross-language communication.

While motivation to be understood is likely not the strong factor in SDA that it is in SLA, motivation to sound like a native speaker of one’s D2 could play an important role in predicting how the pattern of long-term phonetic accommodation will unfold. Acquirers of a D2 who wish to integrate into their new community to be accepted by the speakers of the D2 may
attempt to do so linguistically, while others, who are perhaps less eager to integrate or who wish to retain the linguistic features of their native dialect may be less inclined to acquire the sound system of the new dialect (Giles & Smith 1979). Whether speakers of a given dialect who are transplanted into a new dialect area will be more or less inclined to acquire the D2 may be at least partly attributed to the notion of social identity, which “refers to the part of a person’s self-image based on the characteristics and attitudes of the social group or groups which that person belongs to or aspires to belong to” (Siegel 2010: 106). Citing Escure (1997), Siegel notes that “people may continue to use features of their D1 rather than those of the D2 not because of factors related to their learning capacity but because of factors related to their social identity – i.e. maintenance of the identity they associate with the D1 or avoidance of the identity they associate with the D2” (Siegel 2010: 107).

2.1.2.1 Second-Dialect Acquisition in Adults

Regardless of the motivation for acquiring a second dialect, studies focusing on the acquisition of a D2 sound system have found evidence that adults’ speech can and frequently does change as a result of moving to a new dialect area. For example, Munro, Derwing and Flege (1999)’s study of the acquisition of the Alabaman dialect of English (a Southern United States dialect) by English-speaking Canadians found evidence of the acquisition to some degree of certain aspects of the Alabaman dialect. In that study, 30 speakers participated: 10 English-speaking Canadians who were living in Alabama, 10 English-speaking Canadians living in Canada, and 10 Alabamans living in Alabama. The participants provided speech samples via a picture description task. Ten second snippets of all the recordings were then presented to Canadian listener judges who assigned an accent rating from 1 (very Canadian) to 9 (very American) and to Alabaman judges who assigned an accent rating from 1 (definitely from Alabama) to 9 (definitely not from Alabama). Both sets of judges consistently rated the Canadians who had been living in Alabama with an intermediate accent between the Canadians in Canada and the Alabamans in Alabama, suggesting that at least certain aspects of the Alabaman dialect had been acquired by the transplanted Canadians and that the changes in their speech was perceptible to native speakers of both the D1 and the D2.

Similarly, Evans & Iverson (2007) tested speakers of a Northern British English dialect at four time periods: before the speakers began university in southern England where they would be exposed to Standard Southern British English (SSBE), three months after beginning...
university, and after their first and second years of studying. The results of an acoustic analysis indicated that after exposure to SSBE the participants showed centralization of the Northern vowel /o/ found in the words *bud* and *cud* towards the SSBE /ʌ/. In addition, according to an accent rating of their speech on a scale of 1 (very northern) to 10 (very southern), the participants were rated as more Southern-sounding after spending time at university.

Shockey (1984) investigated how speakers of American English (AE) who emigrate to England might come to suppress flapping of /t/ and /d/, a characteristic of North American English. She tested 4 male speakers who had moved to England as adults and who had lived there for a minimum of 8 years. She investigated the realization of flapping in words where /t/ or /d/ would be unequivocally flapped in AE. The results showed that the Americans resisted flapping /t/ between 50% and 83% of the time, and resisted flapping /d/ between 28% and 42% of the time, indicating that there was variation between the speakers in the extent to which they had lost this characteristic of their D1.

Similarly, Foreman (2000) considered the acquisition of three phonetic variables of Australian English by American English speakers who had all been living in Australia for at least 20 years. The differences between these two dialects that were considered were non-prevocalic /r/, which is explicitly realized in American English, but not Australian English, and differences in realization of two vowels: the vowel /ɪ/, which is higher in Australian English than it is in American English, and /o/, which is typically realized as [ʊ] or [o] in American English and as [ʊo] or [ʊ] in Australian English. Foreman found significant variability between the participants in the extent to which they had acquired the phonetic variables. In her study, Foreman evaluated the production of words containing the phonetic variables and perceptually determined whether each had been produced as the Australian or American English variant. She calculated the degree of acquisition by each participant via the proportion of the tokens containing each phonetic variable that were realized as the Australian variant. Three of the participants showed no evidence of having acquired any of the Australian variants, one participant showed “some limited change” in one of the variables, but not the others, and two of the participants showed “significant” shifts in pronunciation patterns towards the Australian norms (Foreman 2000: 5). As noted by Foreman, each of the dialectal differences has intermediate forms between the canonical Australian and American variants, but each realization was categorized as being either Australian or American without considering
intermediate forms. It is possible that the participants in Foreman’s study had indeed made phonetic changes to their realization of the variables under investigation, but not to the degree that it caused Foreman to perceive the variable as being produced as the Australian variant. To the extent that the participants had made within-category phonetic adjustments to their realizations of these phonetic variables, the degree of acquisition of the Australian variants will be underestimated in Foreman’s study. Foreman also suggests that the participants who identified more strongly as Australian may have been more likely to acquire the Australian variants than those who identify more strongly as American.

For Spanish, Pesqueira (2008) investigated the acquisition of the Mexican Spanish glide /j/ by native speakers of Argentine Spanish who had immigrated to Mexico City. The study focused on lexical items containing orthographic <y>, which is realized as /j/ by speakers of Mexican Spanish, but as the alveopalatal fricative /ʃ/ or its voiced variant [ʒ] by speakers of Argentine Spanish. Pesqueira’s study included 12 participants (6 men and 6 women), all of whom had spent a minimum of two years in Mexico City and who were 27 years of age or older. The methodology was comprised of sociolinguistic interviews that lasted about 45 minutes per participant where the participants’ experiences in Mexico City with the language and culture and any linguistic problems they may have had in communicating with Mexicans were discussed. In addition, the participants completed a thorough sociolinguistic questionnaire, performed a lexical naming task to determine whether they used Mexican or Argentine terms for various items, and read a list of words containing <y> aloud. Pesqueira analyzed the recordings by counting the number of instances <y> was realized as the Mexican glide /j/ and the number of instances that it was realized using the native Argentine fricative /ʃ/. She found that various linguistic and social factors affected the percentage of use of the Mexican variant. The linguistic factors that favoured the use of the Mexican glide were lexical frequency (more frequent words used the Mexican glide more often) and whether or not the word was likely to have been learned in Mexico. It is impossible to know for certain if a particular word had been learned in Mexico or in Argentina, but Pesqueira assumed that words that referred to street names, neighbourhoods, or Mexican food were likely to have been learned after the participants moved to Mexico City. The results showed that words that were learned in Mexico had a probability of 73% of containing the Mexican variant, while those learned in Argentina had only a 49.1% probability. A variety of social factors also affected the proportional use of the Mexican glide. Pesqueira’s results show that being a woman, having a positive attitude towards Mexican
Spanish, having lived many years in Mexico, having little contact with other Argentine Spanish speakers, having a Mexican partner, and having plans to remain living in Mexico all favoured the use of the Mexican glide over the Argentine Spanish variant. Pesqueira’s findings are interesting because they show variation in the extent of accommodation to the Mexican glide between participants, but suggest that this variation has much to do with individual social factors, such as speaker attitudes to the new dialect.

2.1.2.2 Second-Dialect Acquisition in Children

Studies focusing specifically on the acquisition of second dialect features by adults, such as those discussed above, are uncommon in the literature; somewhat more attention has been given to dialect change in children, often focusing on the effect of age of arrival in a new dialect area (e.g. Payne 1980). We might expect that some of the results of dialect acquisition in children will extend to studies of adults, but we might also expect there to be some differences in the course of SDA between the two populations.

Chambers (1992) studied the path of acquiring D2 features and losing D1 features in 6 Canadian youths from two families who moved from Canada to south Oxfordshire, England in 1983 and 1984. In this case the D1 was Canadian English (CE) and the D2 was Southern England English (SEE). Chambers conducted sociolinguistic interviews with the youths at two points two years apart in which the youngsters discussed their experiences in their former and new homes, performed a picture-description task, read a word list aloud, and evaluated accents on tape. With these data, Chambers generated eight principles of dialect acquisition. Two of these principles are directly relevant to the present study and will be discussed here. The first is Chambers’ sixth principle, which states that “phonological innovations are actuated as pronunciation variants” (Chambers 1992: 693). This means that speakers make pronunciation changes on a word-by-word basis rather than implementing a global change in realization of a particular phoneme. As discussed by Chambers, this pattern is in line with the predictions of lexical diffusion (Chen & Wang 1975), a pattern of sound change in which phonological changes become regular by individual instances of the rule or new phoneme being acquired (each within a lexical item) until the quantity of acquired instances reaches a particular critical mass, at which point the rule becomes regular. Trudgill (1986) also discusses this pattern, explaining that it is not the entire phonological system that speakers make changes to, but rather “they modify their pronunciation of particular words, in the first instance, with some words
being affected before others” (Trudgill 1986: 58). Trudgill further notes that the motivation for this pattern of change is to “make individual words sound the same as when they are pronounced by speakers of the target variety” (Trudgill 1986: 58). This principle of dialect acquisition given by Chambers (1992) is important to the present study because it suggests that we might find that the participants converge more on the lexical items that are included in the conversation component of the experiment (described in §5.2.4), and therefore are heard spoken aloud by a speaker of the opposing dialect, than they do on lexical items not present in the experiment. This possibility will be considered in §6.

The second principle of dialect acquisition given by Chambers that concerns the current study is his eighth principle which posits that “orthographically distinct variants are acquired faster than orthographically obscure ones” (Chambers 1992: 697). In essence, this principle predicts that dialect variables that sync with the orthography will be easier to acquire than those that do not. The example he gives to support this is a discrepancy in the acquisition rates of two variables by the Canadian youngsters living in England. The variables are loss of voicing (or flapping) of intervocalic /t/ (a characteristic of North American English) and the acquisition of /r/-lessness (the zero realization of non-prevocalic /r/ in many dialects of English). Chambers notes that in the case of loss of /t/-voicing, the outcome syncs with the orthography in that lexical items that are affected contain the grapheme <t>, which after losing /t/-voicing is realized as the voiceless stop /t/, rather than a voiced allophone. In contrast, acquisition of /r/-lessness does not sync with the orthography since the grapheme <r> is present in words such as forty, but is not pronounced explicitly in British English as it is in Canadian English. According to Chambers, this syncing with the orthography is what causes loss of /t/-voicing to proceed faster in the Canadian youngsters in his study than acquisition of /r/-lessness. Trudgill (1986: 11) also notes that dialect variants that do not sync with the orthography can be stigmatized and, as a result, potentially more salient to speakers. If the extent that the dialectal differences between BAS and MS in the present study sync with the orthography contributes to the perceptual salience of the differences, then this is relevant for the present study. This idea will be considered further in §4 when we discuss the particular six dialectal differences that will be investigated in this thesis.

Although the above studies on the acquisition of a second dialect indicate that both children and adults do make changes to their speech as a result of moving to a new dialect area,
it is likely the case that adults in particular will not completely acquire a D2 (Munro et al. 1999). Even if a D2 learner makes significant changes on a particular linguistic variable towards the norms of the D2, a native speaker of the D2 may still be able to perceive a discrepancy. Similarly, the learner could completely acquire one linguistic variable, but make no changes to another, leaving traces of her D1 in her speech, or have any configuration of degree of acquisition of the D2 variables (Foreman 2000; Bauer 2008; Babel 2010, 2012). For the current study, the focus is not on the level of ultimate attainment in SDA, but rather the likely more subtle changes that take place during a single interaction with a speaker of a D2. There is a small, but growing, body of research considering short-term phonetic accommodation that has focused exclusively on adults, which provides a solid foundation for the present study and will be considered next.

2.2 Short-Term Phonetic Accommodation

As discussed in the previous section, several researchers have suggested that long-term changes in adults’ pronunciation towards a D2 may result from repeated face-to-face interactions with speakers of the D2 whereby the short-term convergences that take place in each interaction accumulate over time (Trudgill 1986: 39; Chambers 1992; Niedzielski & Giles 1996). Research on short-term accommodation has found that two people in an interaction often come to be aligned in various aspects of human behaviour (e.g. Chartrand & Bargh 1999), including different characteristics of speech. The alignment of the acoustic-phonetic properties of speech is referred to as phonetic accommodation, convergence or imitation. This section will summarize the existing literature on short-term phonetic accommodation in terms of the motivations, methods, and findings to make clear the social, situational, and linguistic factors that seem to affect the process. Furthermore, this section will introduce the particular theoretical frameworks that studies of short-term accommodation are typically couched in and discuss their implications for the motivation of and mechanism behind accommodation.

2.2.1 Communication Accommodation Theory

Much of the previous research on accommodation has come from the literatures of social psychology and sociolinguistics. Within social psychology, studies of accommodation are often discussed in the theoretical framework of Communication Accommodation Theory (CAT: Giles 1973), a model of modifications of speech style during an interaction. Under CAT, speakers use
these modifications in speech strategically to create the desired social distance between themselves and their interlocutors. The approximation strategies speakers can use include convergence (becoming more similar to their interlocutor), divergence (becoming less similar to their interlocutor), and maintenance (no change). CAT posits that the main motivation for these accommodations is social in that speakers have a “need (often unconscious) for social integration or identification with another” (Giles et al. 1991: 18) and that speakers will converge in order to decrease social distance between themselves and their interlocutors, to increase intelligibility, or to improve communication. In contrast, speakers are predicted to diverge in order to increase social distance, accentuate distinctiveness, show disdain, or signal group membership (Shepard, Giles & Le Poire 2001).

Central to CAT is the idea that the use of accommodation strategies is highly dependent on the situational context in which two speakers interact, with many variables, such as gender of the speakers and role within the conversation, affecting both the direction of the change in speech (convergence or divergence) and the degree or magnitude of the change. Many of these variables will be discussed in this chapter. With so many factors at play, the pattern of accommodation can be quite complex. One speaker in an interaction may converge, while the other diverges or speakers may converge on one dimension, but diverge on another (Shepard et al. 2001).

The direction and degree of accommodation seem to have consequences for how external judges evaluate both the participants in an interaction and the interaction itself. Studies have found that interactions in which convergence takes place are generally rated more positively than those in which divergence or maintenance occurs (Giles & Smith 1979; Street 1982). In addition, when convergence occurs, the interlocutors are rated as being more likeable and attractive. Chartrand & Bargh (1999) studied behaviour mimicry for movements in conversation such as smiling, face rubbing, foot shaking, and leg crossing. They suggest that “individuals use behaviour mimicry as a communication tool on a completely nonconscious level” (Chartrand & Bargh 1999: 901). Specifically, the authors investigated the extent to which mirroring of behaviours such as those listed above contributes to liking of the mimicker and a general sense of smoothness and success of the interaction. They found that participants whose movements were mirrored by the other participant, who was a research assistant for the study,
judged the conversation to have proceeded more smoothly and liked the other participant more than those in the control group.

2.2.2 Phonetic Imitation in Non-Social Contexts

Phonetic accommodation has also been investigated in the linguistic literature, both in social contexts such as conversation, and in situations more removed from a natural communicative context, such as the rapid shadowing paradigm (Goldinger 1998). In rapid shadowing, participants hear pre-recorded stimuli and repeat them immediately. In these less socially-contextualized situations, much of the research on convergence has considered the role that the relationship between the perception and production systems of speech might play in explaining the mechanism behind accommodation. Some studies have suggested that the process may be less motivated by social concerns, as predicted by CAT, and instead may occur automatically as a result of a direct link between perception and production. Such an account holds that speech is perceived in terms of articulatory gestures that are, in turn, used as parameters for speech production, a relation that results in automatic and unintentional imitation (Goldinger 1998; Pickering & Garrod 2004; Shockley et al. 2004). However, as noted by Krauss & Pardo (2004), the automatic alignment account falters in explaining why divergence in speech patterns is also found, as well as how speakers seem to use speech changes for social purposes.

Laboratory-based studies employing rapid shadowing, in which participants hear recorded stimuli and immediately repeat them, have found that speakers imitate fine-grained acoustic information in the signal, where the stimuli are pre-recorded and presumably there would be no social motivation (such as decreasing social distance) to do so. Goldinger (1998) used shadowing to investigate the imitative tendency across various factors such as number of previous repetitions, duration of delay between presentation of stimulus and repetition, word frequency, and voice identity. Listener judges performed an AXB perceptual task, in which participants hear three repetitions of a particular stimulus and their task is to decide whether the second repetition (X) is more similar to the first (A) or the third (B). In Goldinger’s study, A and B were either a baseline production or a shadowed production and X was a model recording to establish whether listeners could perceive imitation in shadowed tokens. If the listeners perceive that the shadowed tokens are more similar to the model recording than the baseline tokens, this is taken to be evidence of imitation of the model recording having occurred. Goldinger found that, in general, the shadowed tokens were considered better imitations of the model recordings,
subject to differing degrees of the effects of the variables listed above. He also found that imitation was detected less frequently when the shadowers were delayed 3-4 seconds before repeating the stimulus as compared to when they immediately repeated it.

Shockley, Sabadini & Fowler (2004) aimed to replicate the results of Goldinger (1998) by showing that listeners detect imitation in repeated speech and also to determine to what extent voice onset time (VOT) is imitated. They conducted two experiments. The first used the rapid shadowing paradigm in which 8 participants read a list of 80 disyllabic English words beginning with /p,t,k/ and later shadowed pre-recorded productions of the same words. 16 listeners performed an AXB task to determine whether the shadowed tokens were better imitations of the pre-recorded tokens than the baseline reading. Their results supported Goldinger (1998) in that the shadowed tokens were statistically significantly more often selected as better imitations of the model recording than the baseline productions. In the second experiment, the authors tested whether participants would shadow tokens in which the VOT of the initial voiceless stops had been extended by twice its original duration. They found that VOT in the shadowed tokens was statistically significantly longer than in the baseline productions, and follow-up investigations revealed that this was not due to a general lengthening effect across the whole word.

Lab studies using other methodologies besides rapid shadowing have also found evidence of phonetic imitation. In one of the few studies explicitly examining phonetic imitation across dialects, Delvaux & Soquet (2007) tested whether native speakers of Liège or Brussels French, two regional dialects of Belgian French, would be affected by being exposed to ambient speech in the other dialect. The authors focused on differences in mel-frequency cepstral coefficients, formant values, and duration of two vowels in these dialects of French. In the study, participants formed and read aloud sentences in between hearing pre-recorded speakers of the other dialect performing the same task over loudspeakers. The participants were not instructed to imitate or even listen to the voices on the loudspeakers; yet, the findings showed that the participants consistently imitated the characteristics of the opposing dialect’s vowels. The authors propose an account in line with probabilistic phonology (Pierrehumbert 2002), explaining that “the imitative mechanism leaves a memory trace, indicating that it is not just a stimulus-and-response behaviour, but that it is a more sophisticated adaptation to the speech
environment involving learning and concomitant modification in gradient phonetic representations” (Delvaux & Soquet 2007: 171).

Another method of exposing participants to a particular speech sample is via blocked exposure, where participants provide baseline pronunciation samples, then listen to a sample of speech which has been altered in some way or which comes from a particular dialect or speech style, and then provide a post-listening speech sample, which can be compared acoustically or perceptually to the baseline. Using blocked exposure, Nielsen (2011) investigated imitation of VOT in order to establish whether imitation of phonetic characteristics is constrained by phonological knowledge. In her study, Nielsen created stimuli with initial /p/ that had either been extended or reduced in VOT. She found that participants who listened to a model speaker producing the /p/-initial words with altered VOT imitated the lengthened VOT, but not the reduced VOT. Furthermore, the imitative effect of the lengthened VOT was found to generalize to novel /p/-initial words and to /k/-initial words, which were not part of the set of stimuli to which the participants were exposed. Nielsen concluded that the participants imitated the extended VOT but not the reduced because the reduced VOT started to encroach upon the values expected for the voiced stop /b/, threatening the phonemic voicing contrast, whereas the lengthened VOT did not threaten any contrast, ultimately suggesting that accommodation is filtered by phonological linguistic knowledge. In addition, the generalization of the imitative effect to novel /p/-initial words provides evidence for a phonemic level of representation, while the generalization to /k/-initial words suggests a featural level – in this case, the feature [spread glottis].

2.2.3 Factors Affecting Phonetic Imitation

2.2.3.1 Social and Situational Factors

According to Nielsen (2011), that phonetic imitation is constrained by linguistic knowledge means that imitation is not an automatic process as predicted by gestural theories of speech perception, but rather is a process that is mediated by various factors. Other studies have found evidence of these intervening factors as well, including gender, conversational role, racial and regional dialect bias, and attitude towards a model speaker. Pardo (2006) investigated the effects of gender and conversational role on phonetic convergence in spontaneous conversation. Her methodology incorporated the map task, first formulated by the Human Communication
Research Center (HCRC) (Anderson et al. 1991). The map task, which is used in the methodology of this thesis, involves the use of a pair of maps, generated by the researcher, which include various labelled landmarks. A pair of participants each receives a copy of the map. One participant’s map includes a route drawn between two points that passes by the various landmarks. The other participant’s map contains the landmarks, but not the route. The goal of the task is for the two participants to communicate together such that the participant without the route is able to draw the route on his map as faithfully to the original as possible.

The map task was particularly well suited to the goals of Pardo (2006) since different roles can be assigned to the two participants completing the task: giver of instructions, and receiver of instructions. Six same-sex dyads (3 female pairs and 3 male pairs), all speakers of English, participated. Before and after completing the map task all participants also read a list of the map task labels aloud. The pre-conversation word list served as a baseline of pronunciation for the participants while the post-conversation word list allowed comparison with tokens produced before and during the map task. Convergence was assessed via an AXB perceptual task in which 30 listener judges participated. As mentioned earlier, in the AXB task, listeners hear three instances of a particular item and decide whether A or B is more similar to X. In Pardo (2006), X was always a map task label taken from the conversation of one of the participants in the task, while A and B were the same label taken from the partner of that participant, and were produced during the pre-task word list, the post-task word list, or during the task itself. If the listeners perceived that the landmark labels produced during or after the map task were more similar to the conversational partner than the labels produced before the task, this was taken to be evidence of convergence between the participants having occurred.

The results of the perceptual task found that convergence did occur and that its effect accumulated during the map task and persisted into the post task. In addition, there were effects of sex and role. Males were found to converge more than females overall, and givers were found to converge more than receivers. Within the male pairs, receivers converged more to givers, whereas in the female pairs the opposite result was found, with givers converging more to receivers. In terms of the sex effects, the results of Pardo (2006) are at odds with those of Namy, Nygaard & Sauerteig (2002), where a lexical shadowing task was utilized to consider the role of gender in accommodation. In that study the authors found that females converged more to a model speaker than males did. As noted by both Pardo (2006) and Namy et al. (2002), the
differences in convergence between the genders may be partly due to differences in ability to detect convergence and attention to detail, rather than differences in inclination to converge.

Babel (2012) considered phonetic imitation in a shadowing task as well and found that the process is mediated by social factors. In her study, English-speaking participants shadowed 50 words containing one of the five vowels /i, æ, u, o, ɑ/ produced by one of two model speakers: a black male or a white male. The degree of imitation was assessed via an acoustic analysis of the first and second formants of the vowels in a baseline production task, the shadowing task, and a post-shadowing production task. The measure used was DIFFERENCE-IN-DISTANCE which reflected the change in Euclidean distance between a participant and the model speaker from baseline to shadowing for each of the five vowels. The participants also performed an Implicit Association Task (IAT), which was designed to measure implicit racial bias. The results showed that phonetic imitation is socially selective since participants who scored with a pro-black bias in the IAT imitated the black model speaker more than those with less of a pro-black bias, suggesting that imitation is mediated by racial bias.

In a study involving similar methodology, Babel (2010) investigated cross-dialectal accommodation by speakers of New Zealand English (NZE). In that study, NZE participants with a pro-Australia bias were found to be more likely to imitate an Australian English model speaker than those with more of a pro-New Zealand bias (as assessed via IAT), suggesting that bias about a region or regional dialect also mediates phonetic imitation.

On a more specific level, a speaker’s opinion about an individual may also affect the likelihood of imitating that person. Abrego-Collier, Grove, Sonderegger & Yu (2011) investigated whether a positive or negative evaluation of a model speaker would affect degree of imitation of extended VOT. In the experiment, participants performed a baseline production task producing 72 /p,t,k/-initial target words, then listened to a short story in which a narrator recounts his experience on a date under two conditions: the positive condition where all goes well and the negative condition where the date is a disaster. All 72 targets words were included in the story and the VOT of the stops in the target words was extended by 100%. After listening to the story, the participants then produced the target words again and completed a post-task survey in which they evaluated their feelings towards the model speaker on a scale of 1 (very positive) to 7 (very negative). The results found that participants with a more positive attitude
towards the model speaker were more likely to imitate the extended VOT than those with a more negative attitude, suggesting that phonetic imitation is mediated by attitudinal factors.

A related result was found in Pardo, Gibbons, Suppes & Krauss (2012), in which convergence between pairs of college roommates was investigated. Speech samples were collected from 5 pairs of college roommates at 4 points during an academic year. Convergence was assessed via an AXB perceptual task. In addition, the researchers used a questionnaire designed to determine the quality of the relationship between the roommates in order to explore how the relationship might affect the pattern of convergence. All 5 pairs were found to converge by the fourth time period and a significant correlation between convergence and reported closeness between roommates was also found, suggesting that roommates who have a closer relationship will converge more.

### 2.2.3.2 Linguistic Factors

The majority of the factors that have been found to affect the process of phonetic convergence that are discussed above are social (gender, racial bias, regional dialect bias, attitude towards model speaker, closeness between conversational partners) or situational (conversational role), but linguistic factors may also play a role. In addition to the finding that implicit racial bias affected phonetic imitation, Babel (2012) found that imitation of English vowels was phonetically selective since it was much more common and greater for the vowels /æ/ and /ɑ/ than the others (/i, u, o/), particularly in the first formant (F1). Babel proposes that this is due to the greater inherent variability of these two vowels resulting from differences in realization according to whether the vowel is prosodically accented or not. Babel explains that changes in jaw height are associated with changes in F1 and that the spectral differences between accented and unaccented vowels have been shown to be greater for low vowels than for high vowels, citing Summers (1987), Beckman, Edwards & Fletcher (1992), and de Jong (1995). She argues that since the low vowels /æ/ and /ɑ/ are more variable with respect to F1, this means that speakers will have more possible production variants along this dimension for these vowels. Babel further argues that her results suggest that imitation is constrained by the pre-existing phonetic repertoire of any given variable for each participant. If the participants have a larger number of production targets for /æ/ and /ɑ/ then they have more opportunity to converge on those vowels since the outcome of converging still falls within their regular production habits.
As discussed earlier, Nielsen (2011) found that imitation is constrained by linguistic knowledge in the form of phonological contrast, a linguistic factor. Another study that focused on the effects of a linguistic factor in the pattern of convergence is Kim, Horton & Bradlow (2011). In that study, convergence was investigated in spontaneous conversation and the purpose was to examine the effect of ‘language distance’: the magnitude of the difference between the speakers’ experience with the language in use. Language distance can be thought of as occupying a continuum from close (two speakers of the same dialect of the same first language) to far (two speakers of different L1s both using their L2); falling in between would be two speakers of different dialects of the same L1. Kim, Horton and Bradlow’s study examined convergence in conversations between native English speakers (two pairs who spoke the same dialect and two pairs who spoke different dialects), between native Korean speakers (two pairs who spoke the same dialect and two pairs who spoke different dialects), and between native English speakers and L1-Korean L2-English speakers or L1-Chinese L2-English speakers (four pairs of each). The conversations came from the Wildcat Corpus and were elicited using the Diapix task (Van Engen, Baese-Berk, Baker, Choi, Kim & Bradlow 2010), in which each talker in a pair is given a picture of a scene, and the pictures are identical except for ten differences. The task is for the talkers to communicate with each other in order to identify the differences. Phonetic convergence was assessed via an XAB perception test, where X was a sample of one talker’s speech and where A and B were samples of the other talker’s speech at either early or late points in the conversation. The results showed that the degree and direction of adjustment of the native participants in relation to the L2-English speakers varied as a function of proficiency, in terms of accentedness of the L2 speakers. Accentedness was evaluated by 15 native English judges by listening to speech samples of the L2 speakers and rating the samples on a scale of 1 (native-accented) to 9 (heavily foreign-accented). The native speakers paired with heavily-accented Korean speakers showed significant divergence from their partners, while those with moderately-accented partners showed maintenance, and one native speaker paired with a highly proficient Korean speaker showed convergence towards his partner.

Building on the notion of language distance, Kim, Horton & Bradlow (2011) also investigated convergence patterns in conversation between two native speakers of the same dialect of either English or Korean and two native speakers of two different dialects of English and Korean. They found that convergence was greatest in the same-dialect pairs and that maintenance and divergence were common in the different-dialect pairs, lending support to the
idea that the closer the language distance, the more likely convergence will occur. Furthermore, the authors found that the dialect effect was greater for the Korean pairs than for the English pairs and suggest that this could be due to the two Korean dialects being more dissimilar than the two English dialects. If convergence is less likely between dialects that have larger differences than between those with smaller differences, this further supports the result that convergence depends on language distance.

Incorporating this notion of language distance, Evans, Alshangiti, Hazan, Baker & Cyrus (in preparation) investigated phonetic convergence in conversation between speakers of two different dialects of English: Northeast English (NE) and Standard Southern British English (SSBE). In that study 6 pairs of female speakers who were friends prior to the experiment participated. The NE speakers were born in northeast England, but moved to London in the south of England for university at 18 years of age and had been living in London for a minimum of 2 years at the time of the experiment. All the participants were students at the University of London. Conversation was elicited via a Diapix task (Van Engen et al. 2010), as in Kim et al. (2011), along with a pre- and post-conversation reading task in which the participants read a phonetically-balanced passage. The Diapix task included 8 key words that contained 4 phonetic differences between NE and SSBE. Convergence was assessed via accent ratings (where 1 = very SSBE, 10 = very NE). Six snippets per speaker were taken from within the Diapix conversation (3 were early, taken from the first third of the conversation, and 3 were late, taken from the last third of the conversation). Within early and late snippets, two were ‘accent revealing’ (AR) in that they contained one of the four phonetic differences between the two dialects, and the other snippet was ‘accent neutral’ (AN) since it did not contain any variable expected to be strongly associated with either dialect. Listeners also gave accent ratings on a short section of the pre- and post-conversation passage.

The hypothesis was that the NE speakers would converge towards SSBE since NE is considered lower prestige than SSBE and that the AN snippets would show more convergence than the AR snippets. Since the AN snippets do not include the phonetic variables purported to be more salient, the authors predict that the less salient dialectal differences would show more convergence than the more salient differences, as found in Babel (2010) and Kim et al. (2011), but contra Trudgill (1986: 11), which will be discussed in detail in §3. The results showed that convergence seemed to be limited to the NE speakers, with no evidence of convergence in the
SSBE speakers. Some NE speakers seemed to diverge from their SSBE interlocutors in the AR snippets (i.e. on the most salient dialectal differences) and two of the NE speakers converged to their SSBE interlocutors on the AN snippets. The authors note that the limited finding of convergence was somewhat of a surprise since robust convergence was found in other studies (such as Pardo 2006), but that the results of Kim et al. (2011) that greater language distance leads to less convergence may support their finding since the pairs of speakers in Evans et al. (in preparation) spoke two different dialects of English. Furthermore, the effect of snippet type suggests that speakers may diverge more on the more salient differences. As suggested by the authors, listeners may also weight the salient differences differently from the less salient differences in making their judgements about accent.

As we have seen, the body of literature investigating phonetic convergence is not extensive, but is growing and much of the research coming from the phonetic literature is recent. Phonetic convergence has been shown to be affected by various social, situational, and linguistic factors, but the majority of these studies have considered the pattern of accommodation between speakers of the same dialect, and almost exclusively speakers of English. Studies examining cross-dialectal convergence are few, and those considering convergence in languages other than English are practically non-existent. A handful of these studies have suggested that differences between two dialects that differ in how perceptually salient they are might not be subject to convergence in the same way. That is, the perceptual salience of these dialectal differences might affect whether and how participants in an interaction converge on them, both in the short term as well as in longer-term patterns of acquisition of a second dialect. The purpose of this thesis is to explore the role that perceptual salience plays in phonetic accommodation in conversation between speakers of two different dialects of Spanish. The following section considers how the notion of perceptual salience has been defined previously in the research, and makes a case for the use of an empirical, context-dependent, individually-based measure.
Chapter 3
Perceptual Salience

As discussed in §2, although a few studies of phonetic accommodation have suggested that perceptual salience might affect the pattern (Babel 2010; Kim, Horton & Bradlow 2011; Evans et al. in preparation), none of these studies has expressly set out to investigate the role of perceptual salience in accommodation, and as a result, none has provided a solid definition of what salience means. Defining the concept of perceptual salience seems intuitive at first glance, but becomes more difficult as we dig deeper into what we really mean by the idea. Kerswill & Williams (2002) term salience “the property of a linguistic item or feature that makes it in some way perceptually and cognitively prominent” (Kerswill & Williams 2002: 81). For Siegel (2010), salience “refers to the characteristic of being easily noticeable, prominent or conspicuous” (Siegel 2010: 129). Hickey (2000) states that “salience is a reference to the degree to which speakers are aware of some linguistic feature” (Hickey 2000: 57). These definitions suggest that a particular linguistic variable is salient if speakers notice it or are aware of it, but they make no prediction about why a particular variable should be more noticeable than another.

Several authors have suggested that the why of salience has two potential sources: inherent salience and derived salience (Winters 2000; Goldscheider & DeKeyser 2001; Yip 2006; Ghia 2011). Ghia (2011) states that “salience may either indicate the inherent relevance of a linguistic feature, or else the prominence acquired by a linguistic item during its delivery as a result of its distinctiveness from input or against previously formed cognitive representations” (Ghia 2011: 2, emphasis in original). A particular linguistic variable may be inherently salient, for example, as a result of being based on an acoustic parameter that is more salient to the human perception system than another. For instance, Bohn (1995) suggests that duration as a cue to vowel contrasts may be inherently more salient than spectral cues since in his study he found that Spanish and Mandarin speakers relied almost exclusively on duration to perceive English tense/lax vowel contrasts, even though duration is not an acoustic correlate of vowel contrasts in either Spanish or Mandarin. Derived salience, on the other hand, does not come from within a linguistic variable, but rather is bestowed upon it by virtue of being in a particular context. This context could be phonological, prosodic, syntactic, or social (among potential others) and when a linguistic variable appears in such a context its salience can be enhanced, but this increased salience is the result of the context, not an inherent property of the variable.
Certainly, we expect interaction between these two conceptualizations of salience. Strange & Shafer (2008) note that much more work is required in determining the inherent salience of various acoustic-phonetic parameters as well as teasing apart inherent salience and derived salience.

Even though the notion of salience remains defined imprecisely in the literature as a whole, it is frequently invoked to explain diverse linguistic phenomena from areas such as first and second language acquisition and within-language perception patterns. For example, Bardovi-Harlig (1987) incorporates salience alongside markedness to explain the order of acquisition of syntactic constructions of English by L2 learners, where salience is defined through the availability of input. Goldschneider & DeKeyser (2001) also make use of salience in their account of the order of L2 acquisition of English morphemes, defining salience via inherent properties of the morpheme, where more salient morphemes are longer, have a vowel in their surface form, or are more sonorous. Within phonology, Hancin-Bhatt’s Feature Competition Model (Hancin-Bhatt 1994) is designed to function as a metric to predict the prominence of various features, which in turn affects how perceptually salient the features are. This salience is then used to explain patterns of segmental transfer during acquisition of an L2. In the model, prominence of a given feature is not defined via listener judgements or another experimental technique, but rather is calculated as the ratio of the total number of phonemes for which that feature is specified to the total number of phonemes in the system.

Perceptual salience has also been incorporated into discussions of adult L1 within-language perception patterns. For example, Chapman (1995) uses salience in her discussion of the productivity of inflectional and derivational vowel alternations in Swiss German, where salience is defined by linguistic factors such as the semantic relation between a vowel alternant and meaning. Hume (1998) uses salience to explain phonological patterns of consonant/consonant metathesis by proposing that certain stop consonants are more perceptually salient than others, and that perceptibility can be enhanced in prosodically prominent contexts. Steriade (2001) proposes the P-Map, which represents how perceptual salience of contrasts can change depending on the prosodic context. Yip (2006) proposes that English consonants vary in their perceptual salience in her account of consonant deletion in English loanwords into Cantonese. Winters (2000) used perceptual salience to explain the context-dependent relative
cue weighting of the acoustic correlates of the voicing contrast of syllable-final stops. Toro, Sebastián-Gallés and Mattys (2009) propose that salience also plays a role in the segmentation of connected speech. This brief review shows that the notion of salience is frequently appealed to in order to explain diverse linguistic phenomena and that its definition can be very distinct from one account to another.

One of the main difficulties associated with studying the role of salience in virtually any linguistic phenomenon is the task of establishing just how salient a particular variable is and whether it is more or less salient than another. Torbert (2004) notes that “every sociolinguist knows what salience means and possesses some notion of which linguistic variables are highly salient, but such notions remain under-investigated and largely un-reflected upon” (Torbert 2004: 2). Hickey (2000) points out that salience is “a phenomenon which is generally recognized by linguists but which is notoriously difficult to quantify” (Hickey 2000: 57). The studies mentioned above have incorporated the notion of perceptual salience in a variety of ways, a handful using experimental methods to explicitly measure speakers’ perceptions, while the majority have relied on structural patterns to predict salience. This divide further complicates the definition of salience: salience can be inherent or derived, and can be explicitly measured or assumed to stem from a set of structural patterns in a given linguistic system.

In this chapter, I will review the studies of dialect acquisition and change that have explicitly incorporated salience, delineating those that attempt to quantify salience as perceived by language users from the others that do not. I will attempt to make the case that in order to understand the role of perceptual salience in accommodation we must measure speakers’ perceptions of the salience of the linguistic variables we are interested in directly, rather than relying on a set of criteria that are assumed to reflect all speakers’ perceptions accurately and equally.

In §3.1, I will discuss the method of defining salience via a list of criteria that are used as a diagnostic for salience, which is commonly employed in studies of dialect acquisition and change. To some extent I will give special focus to the definition of salience and the investigation of its effects provided in Trudgill’s book *Dialects in Contact* (Trudgill 1986) since it is one of the original and most influential formulations of salience that considers its role in long-term dialect accommodation. In §3.2, I will evaluate the criteria-list approach to salience
and examine its limitations and problems. In §3.3, I will discuss studies that have used experimental methods to explicitly measure speakers’ perceptions of linguistic variables in order to quantify perceptual salience and attempt to make a case for the experimental method. In §3.4, I will consider the role of experimental methods in defining salience within studies of phonetic accommodation and propose the definition of salience that will be used in the remainder of this thesis.

3.1 Salience as a List of Criteria

3.1.1 Salience in Short- and Long-Term Accommodation

As noted earlier, the definitions of salience that have been proposed in the literature on dialect acquisition and change often consist of a list of criteria that a given variant will meet in order to be considered salient. The most frequently cited and applied definition of salience comes from Trudgill (1986)’s prominent work on dialect contact. Trudgill’s criteria for salience have been applied in several studies by other researchers, mostly focusing on long-term dialect acquisition and change. Some of his criteria have found support, while others have not or have been criticized for being based on circular logic, as will be discussed below.

In his discussion of accommodation between different dialects, Trudgill (1986) states that “in contact with speakers of other language varieties, speakers modify those features of their own varieties of which they are most aware” (Trudgill 1986: 11). This seems to suggest that in interactions between speakers of two different dialects, it is the variants of the individual speaker’s own dialect of which he or she is aware, rather than the variants of the other dialect. However, Trudgill also talks about accommodation occurring “by the modification of those aspects of the segmental phonology that are salient in the accent to be accommodated to” (Trudgill 1986: 20), suggesting that it is the salience of the second dialect variants that speakers are sensitive to; but, as discussed by Siegel (2010: 120), awareness or salience likely results from noticing a contrast or difference between the realizations of a linguistic variable between two dialects, rather than the inherent salience of a particular sound itself. For example, in Madrid Spanish, orthographic <c> (before <e> or <i>) and <z> are realized as the voiceless interdental fricative /θ/. This fricative in and of itself is not likely to be inherently more salient for speakers of this dialect than any of the other consonants in the inventory, which are all in phonological contrast with each other; but, when the realization of orthographic <c> (before <e>
or \(<i>\) and \(<z>\) in MS is compared to that of BAS (where these graphemes are produced as /s/), the difference could potentially become salient for speakers of both dialects; the speakers may notice and become aware that the realizations are different.

After proposing that speakers modify the differences between two dialects of which they are most aware in cross-dialectal interactions, Trudgill (1986: 11) further proposes that there are four characteristics of linguistic variables that will cause speakers to be more aware of them than they are of other variables. These four criteria are given in Table 1.

Table 1: *Trudgill (1986)*’s four criteria for salience of a linguistic variable

<table>
<thead>
<tr>
<th>#</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Speakers will be more aware of forms that have a variant that is overtly stigmatized</td>
</tr>
<tr>
<td>2</td>
<td>Speakers will be more aware of forms that are involved in an ongoing linguistic change</td>
</tr>
<tr>
<td>3</td>
<td>Speakers will be more aware of forms whose variants are very different phonetically</td>
</tr>
<tr>
<td>4</td>
<td>Speakers will be more aware of forms that are involved in maintenance of phonological contrast</td>
</tr>
</tbody>
</table>

The first criterion states that a variant will be salient when it is stigmatized for being considered too rural or uneducated. Stigmatized variants are often included in mimicking or jokes about speakers of a dialect that contains those variants (Hickey 2000; Siegel 2010: 120); however, it is not clear why stigmatized variants should be expected to be accommodated to. Trudgill’s prediction is that stigmatized variants should be salient, and that the salient variants are those that speakers will accommodate to, but stigmatized variants can hardly be expected to be willingly incorporated into one’s speech, except in cases where such variants carry covert prestige (Trudgill 1972; Sorenson 2011). As noted by Siegel (2010: 121), a more likely prediction is that a speaker would be more apt to acquire a D2 variant if the corresponding D1 variant were stigmatized. Although not explicitly stated, it is possible that this is the direction of accommodation that Trudgill intended as well.

Trudgill’s second criterion for salience is that a variable will be considered salient if it is undergoing linguistic change. Bowie (2000) found support for this idea in his study of Waldorf, Maryland exiles, stating that “those things that are not already in a state of change in the native dialect appear more resistant to change upon constant exposure to a new dialect than those things that are in a state of change” (Bowie 2000: 125). He found that the vowels of the Waldorf
English dialect that were most stable, and therefore least salient according to Trudgill’s second
criterion, such as /ʌ/, which Bowie notes had not changed in Waldorf for at least a century,
showed almost no change in the speech of the Waldorf exiles who had moved to other dialect
areas. In contrast, the vowel /æ/, which Bowie notes is undergoing a linguistic change, was more
susceptible to change due to exposure to other dialects. Bowie’s study supports Trudgill’s
proposal that variables undergoing change would be more subject to accommodation, but Bowie
does not explicitly state that these variables would be more salient than others. On the other
hand, in Payne’s (1980) study of the second-dialect acquisition of various aspects of the sound
system of the Philadelphia dialect of English by children from other dialect areas, she concluded
that the variables that were undergoing a change in her study were considered less salient by the
participants. Thus, these studies provide contradictory evidence for whether variables
undergoing a linguistic change would be more or less salient or subject to accommodation.

Trudgill’s third criterion for salience states that linguistic variables whose variants are
“phonetically radically different” (Trudgill 1986: 11) will be considered salient. Presumably this
is because very different variants would be easily noticed. Defining phonetic distance between
two variants is a difficult task; phonetic distance (or similarity) may reflect articulatory,
acoustic, or perceptual similarity (Mielke 2012). In the literature, phonetic similarity has been
commonly quantified using acoustic measurements, particularly for vowels, which have been
shown to be reasonably well represented in a two-dimensional acoustic space using the first and
second formants (F1-F2 space). While acoustic measurement allows us to quantify similarity to
some degree, there are some difficulties with this method. The main problem is that differences
between vowels within the F1-F2 space may stem from factors other than phonemic difference.
For example, variation due to coarticulation, formality, and prosodic and phonological patterns
along with interspeaker variability from physiological differences such as vocal tract size all
may cause a difference between two vowels in the F1-F2 space. In addition, the F1-F2 space
does not capture other acoustic measures that may be exploited as cues by listeners, such as
fundamental frequency, duration or formant movement (Flege, Bohn & Jang 1997).

When listeners use different cues beyond the first and second formants to perceive
similarity between vowels, phonetic similarity as determined via F1 and F2 does not align with
perceptual similarity, or how the listeners actually perceive the similarity between two vowels.
For example, Cebrian (2002) considered the perception of English vowels by Catalan speakers.
According to his acoustic analysis utilizing the F1-F2 space, the three pairs of English-Catalan vowels he considered (/i/-/i/, /ɪ/-/e/, and /ɛ/-/ɛ/) had comparable degrees of similarity. For example, English /i/ and Catalan /i/ were considered as phonetically similar to each other as English /ɪ/ and Catalan /e/ were. However, the results of a vowel identification and category goodness judgement task found that although the phonetic similarity of these three pairs was comparable, Catalan listeners perceived the /ɪ/-/e/ pair to be less similar than the other two pairs, showing that the perceptual similarity did not always fall in line with the acoustic similarity.

Levy & Strange (2008) also found evidence of a distinction between phonetic and perceptual similarity in their study of the discrimination of Parisian French vowel pairs by English speakers who were either inexperienced or experienced with French. Although spectrally the French vowel pair /i/-/y/ is much more similar than the pair /y/-/u/, the participants in Levy and Strange’s study nevertheless were much more accurate at discriminating between the former pair than the latter, indicating that spectral similarity is not the only factor at play in determining how well listeners can distinguish between vowels. These studies suggest that perceptual similarity cannot be deduced solely from phonetic similarity (or distance), and that perceptual similarity must be tested empirically in a perceptual discrimination or identification task (Flege, Bohn & Jang 1997).

The final criterion for salience that Trudgill (1986: 11) proposes is that speakers will be more aware of forms that are involved in the maintenance of a phonological contrast. Although this criterion seems logical – speakers will most certainly be aware, at least at some level, of sounds that signal a change in lexical meaning, i.e. those that are contrastive – it is not immediately clear how this criterion should be applied. One possibility is that if there is a phonological contrast in the D1 which is neutralized in the D2 or vice versa, then it is the difference in the status of the contrast as phonemic (or not) that matters, with this difference making the variants in the D2 salient for acquirers. Another possibility is that it is the status of a pair of sounds as being in phonological contrast in only the D1 that matters, with any phonological contrast in the D1 being salient, such that, for example, the /ɑ/-/ɔ/ contrast in dialects of English that maintain this contrast would be salient for speakers of that dialect because it is contrastive for them. Or perhaps it is the status of a pair of sounds as being in contrast in the D2 that matters. For example, speakers of dialects of English that do not maintain the /ɑ/-/ɔ/ contrast might find the difference between these two sounds salient because they are
contrastive in a different dialect. It seems that it must be a comparison of the two systems that is relevant. As noted by Auer, Barden & Grosskopf (1998), “what constitutes a phonological merger in the acquisition of variety A by speakers of variety B constitutes a phonological split in the opposite acquisition process” (Auer et al. 1998: 166).

While the status of a pair of variants as being in phonological contrast or not may contribute to the variants being considered salient, it is not clear that this status would result in those variants being accommodated to. If a D1 speaker with a phonological contrast between two segments is acquiring a D2 with one merged category between those two segments, the acquirer will likely face instances of homonymic clash during acquisition. For example, in Payne (1980)’s study of the acquisition of the Philadelphia English dialect, there were two variables that were merged in the D2 that were in contrast in the acquirers’ D1: pairs such as *ferry/furry* are both produced with the vowel /ər/, and pairs such as *sure/shore* are both produced with the vowel /ʌr/ in Philadelphia English. While mergers are typically acquired more easily than splits (Auer et al. 1998), a merger that causes homonymic clash for the acquirer may be acquired more slowly than one that does not. Payne found that these two variables were acquired more slowly than others, although the difference between the status of the segments in the D1 and the D2 was likely very salient, since it caused a loss of lexical contrast for the acquirers.

To summarize, Trudgill (1986: 11) proposes four criteria for linguistic variables to be salient and predicts that speakers, in an interaction with a speaker of a different dialect, will accommodate to those variables in the D2 that are most salient. While in most cases these criteria make intuitive sense, studies that have considered their role in acquisition of a second dialect have often been at odds with each other, casting doubt on whether Trudgill’s criteria actually do indicate salience of a variant, or whether it is indeed the more salient variables that should be expected to be accommodated to. As Auer et al. (1998) note, “it is not a priori clear why a feature…should be given up or adopted simply because it is salient” (Auer at al. 1998: 167). In fact, some of the D2 variants investigated in the studies discussed above that would be considered salient according to Trudgill’s criteria were not those that were accommodated to, while others that should have been less salient were accommodated to more. To deal with some of these inconsistencies in patterning, Trudgill proposes three overriding factors that could cause a linguistic variable not to be acquired in long-term dialect accommodation (Trudgill 1986: 17-
The first is the existence of homonymic clash (discussed above), which would cause a loss of lexical contrast. The second factor is the presence of a phonotactic constraint that disallows a structure in the D2. Trudgill makes use of this factor in his discussion of the lack of accommodation of British English (BE) speakers to non-prevocalic /r/ in American English (AE). Since there is no phonetic similarity between the variants (/r/ in AE versus the absence of a segment in BE), this dialectal difference would be predicted to be highly salient, yet BE speakers do not accommodate to it. Trudgill proposes that the overriding factor in this case is a phonotactic constraint that prohibits production of /r/ in non-prevocalic position. The third factor that Trudgill proposes could preclude accommodation to salient variables is if a variant is so salient that it projects a very strong stereotype of the D2, speakers of the D1 may be less likely to take on the D2 realization (Trudgill 1986: 18).

3.1.2 Objective and Subjective Criteria for Salience

As mentioned earlier, other researchers have incorporated some or all of Trudgill’s criteria for salience in their studies of long-term accommodation patterns such as the acquisition of a second dialect by individuals or the changes that take place at the community level in situations of dialect contact. Focusing on the first of these two patterns, Auer, Barden & Grosskopf (1998) discuss the role of salience in predicting which variants of a D1 will be lost first during acquisition of a second dialect of German. They review the various definitions of salience that had been proposed in the sociolinguistic literature by other authors (such as Schirmunski 1930 and Hinskens 1996), including Trudgill’s criteria. The authors distinguish between objective and subjective criteria for salience. Objective criteria are based on a description of the structural properties of a dialect, such as the status of a particular difference as phonemic versus non-phonemic, and do not reflect the speakers’ actual perceptions. In contrast, subjective criteria for salience reflect how speakers perceive the differences between two dialects, such as the perceptual distance between a D1 variant and a D2 variant. The objective and subjective criteria for salience that Auer and colleagues consider in their study are listed in Table 2.
Table 2: *Objective and subjective criteria for salience* (Auer et al. 1998)

<table>
<thead>
<tr>
<th>Objective criteria</th>
<th>Subjective criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>articulatory distance</td>
<td>perceptual distance</td>
</tr>
<tr>
<td>areal distribution</td>
<td>usage in code-alternation</td>
</tr>
<tr>
<td>phonemicity</td>
<td>representation in lay dialect writing</td>
</tr>
<tr>
<td>continuous vs. dichotomous</td>
<td>stereotyping/mimicking</td>
</tr>
<tr>
<td>lexicalization</td>
<td>comprehensibility</td>
</tr>
</tbody>
</table>

Here we see some overlap between Trudgill’s criteria and those proposed by other authors in both the objective and the subjective criteria. Within the objective criteria, *articulatory distance*, proposed by Schirmunski (1930), is the same as Trudgill’s *phonetic distance* criterion. It is expected that the greater the distance, the more salient the difference between the variants will be. Schirmunski also proposes that *phonemicity* (equivalent to Trudgill’s *phonological contrast*) will cause a variant to be salient. *Areal distribution* (Hinskens 1996) refers to how widespread a given dialectal variant is geographically, with more widespread variants predicted to be more salient (a somewhat unexpected prediction). The *continuous versus dichotomous* criterion suggests that linguistic variables that have only two variants with no intermediate forms (no gradient realization), such as the realization of non-prevocalic /r/ in rhotic dialects of English or its absence in non-rhotic dialects, are more salient than those that have a gradient realization such as variations in vowel quality (Schirmunski 1930). The *lexicalization* criterion refers to whether or not the application of a phonological rule depends on the particular lexical item that it is to apply in. Phonological rules that are lexicalized and only apply in specific words are predicted to be more salient than those that apply uniformly throughout the lexicon (Schirmunski 1930).

Within the subjective criteria, the greater the *perceptual distance* between a D1 and a D2 variant, the greater the salience, as proposed by Schirmunski (1930). This criterion differs from its objective counterpart (*articulatory distance*) in that it reflects how speakers perceive the difference, rather than the magnitude of that distance along some measurable dimension, such as F1-F2 space as discussed in the previous section. Variables that are more perceptually distant are predicted to be more salient. *Usage in code-alternation* refers to the presence of a given D2 variant when a bi-dialectal speaker is talking in a situation where code-switching between the two dialects occurs. If the D2 variant is used in code-alternation it is predicted to be salient. Similarly, if the orthographic form of a variant is present in writings by non-linguists (*representation in lay-dialect writing*), that variant is also predicted to be salient. Overlapping
with Trudgill’s *stigmatization* criterion, Schirmunski’s *stereotyping/mimicking* criterion suggests that those variants that are included when imitating a dialect or as a part of jokes involving that dialect will be salient. Lastly, if the difference between a D1 and a D2 variant is likely to hinder *comprehensibility*, then the difference will be salient.

Auer et al. (1998) applied both the objective and subjective criteria for salience given in Table 2 to an examination of a longitudinal study of dialect accommodation of 56 speakers of Upper Saxonian Vernacular (USV) German (which is only spoken in East Germany) who move to other dialect areas where either Upper German or Middle German is spoken. Specifically, they considered the degree of loss of the USV features (rather than the acquisition of the D2 features) over a 2-year period. In examinations of long-term dialect accommodation, a distinction between dialect loss and dialect acquisition is crucial since in applying the predictions of the criteria for salience, the direction of the change matters. For example, Schirmunski’s criterion of *lexicalization* predicts that a phonological rule that is limited to specific lexical items will be more salient and therefore, more learnable and acquired/lost earlier (Auer et al. 1998). However, in Chambers (1992)’s study of the acquisition of British English by Canadian youngsters, it was the lexicalized rules that were acquired later, not earlier. This contradiction results from the application of the salience of the rule equally whether a learner is giving up the rule or acquiring it. If a phonological rule is lexicalized and therefore salient, giving it up (loss of the D1 variant) is predicted to occur early, while in contrast, acquisition of such a rule as a D2 variant is predicted to be more difficult and occur later, if at all. According to Auer et al. (1998), “accommodating speakers may refrain from using such a ‘difficult’ variable at all” (Auer et al. 1998: 168).

The results of Auer et al. (1998) found that the subjective criteria for salience were more effective than objective criteria in accounting for which variants of USV (the D1) would be lost first, at least in the case of the variables that had intermediate forms\(^1\). The four variables that were considered to be low in salience as determined by the subjective criteria were lost at rates of -7%, 17%, 18%, and 19%. The two variables with medium salience were lost at rates of 25% and 39%, while the three variables that were highly salient were lost at rates of 26%, 30%, and 47%. This result suggests that it is the more salient D1 variants that are lost first.

---

\(^1\) Variables with intermediate forms refers to linguistic variables with gradient realizations between the D1 and D2 norms, rather than one D1 variant and one D2 variant.
It is perhaps not surprising that subjective criteria that reflect speakers’ perceptions of the differences between two dialects would be more effective at accounting for patterns of dialect accommodation than objective criteria. For example, that a given D2 variant is articulatorily or acoustically very different from its D1 counterpart (an objective criterion) will only be relevant in accommodation if the speakers involved in accommodating between the two dialects perceive the difference as large (a subjective criterion). How the speakers would perceive the difference depends on the language and variable in question, and the speakers’ experience with the language (e.g. Narayan 2008). Likewise, if a variant’s areal spread is large with that variant being used in many geographical areas (an objective criterion), it is not necessarily the case that a speaker will have such knowledge, even though the spread could be described by linguists.

Although the subjective criteria did correlate reasonably well with the pattern of loss of D1 variants, they were still applied at the group level and assumed to reflect the perceived salience of the D1-D2 differences equally for all of the participants who took part in Auer et al.’s study. The perceptions of the actual participants in the study were not tested. This assumes that the participants’ perception patterns (had they been tested) would have made the same predictions for salience that the subjective criteria given in Table 2 did. Without considering the individual-level perception pattern, potentially explanatory patterns at the speaker level may be obscured. This will be discussed further in §3.2.

3.1.3 Extra-Linguistic Factors for Salience

Further applying a Trudgill-style criteria list for salience, Kerswill & Williams (2002) examined how salience affects community-level sound changes. The authors set out to determine not only the role of salience in sound change through dialect contact, but also to wade through the different criteria for salience that had been previously proposed and establish which criteria are essential to a functional definition of the notion.

Kerswill and Williams suggest that in all of the studies incorporating salience that they review they can “detect an element of circularity” (Kerswill & Williams 2002: 85), noting that previous authors claim to explain the patterning of a particular linguistic variable simply by labelling it as ‘salient’. Hinskens (1996) and Kerswill & Williams (2002) discuss the idea given in Trudgill (1986) that more salient dialectal variants will be accommodated to, but that variants that have extra-strong salience will not be accommodated to, and suggest that such a conception
is circular since “salience is sometimes used as an explanation for accommodation…and sometimes to explain why accommodation does not take place” (Hinskens 1996: 11, emphasis in original). According to Kerswill and Williams, in order to avoid circularity, the definition of salience must make reference to the effects of extra-linguistic factors, such as social relations between the two dialect communities in contact, time scale, intensity of contact, and the involvement of adults or children in the contact situation. The purpose of their study was to investigate dialect levelling between three communities in England in order to determine the roles of language-internal factors, such as phonetic distance between two variants of a linguistic variable that is being levelled or involvement of the variable in a phonological contrast (two examples of the criteria for salience proposed by Trudgill 1986: 11), and extra-linguistic factors, such as those listed above.

The study considered the community-level convergence of linguistic variables, or dialect levelling, across urban centres in England, including three towns: Milton Keynes, Reading, and Hull. The authors note that Milton Keynes, located 70km northwest of London, England, has a very mobile population. Reading is 60km west of London and is “well-established”, while Hull, a town in the north has industries that are declining in prosperity. These particular towns were chosen in order to study the spread of dialect features between North and South towns (Hull versus Milton Keynes and Reading) and between mobile and stable communities (Milton Keynes versus Reading).

The study involved recording the speech of 32 teenagers (14-15 years old) and 4 elderly working-class residents from each town in interviews and discussions in groups. Various linguistic variables were investigated including 8 phonological variables. Four of these were the realization of the vowels in the lexical sets PRICE, GOOSE, GOAT and MOUTH, and four were consonantal variables: [t]-glottaling, fronting of [θ] to [f], fronting of [ð] to [v], and [h]-dropping. The results show that the consonantal features spread from the South to the North, but found no evidence of convergence of vowels between the dialects. Kerswill and Williams attempted to explain the social and geographical spread of these linguistic variables between the three communities in their study by appealing to markedness and salience. The results of their analysis of the eight phonological variables cause the authors to conclude that the notion of salience is only useful if it incorporates extra-linguistic factors and that even though there might be a link between the language-internal factors proposed by Trudgill and extra-linguistic factors,
it is the latter that actually motivate speakers to behave in a particular way, making extra-linguistic factors essential to a functional definition of salience.

3.2 Evaluation of Criteria-List Definition of Salience

The researchers whose studies are summarized in the preceding sections reasoned that perceptual salience may play an important role in accounting for accommodation patterns. They generated working definitions of salience and applied the definitions to real world data. The definitions were based on lists of criteria that were predicted to make a given linguistic variable salient and these lists were incorporated into studies of accommodation to a second dialect to try and explain the order and degree of the acquisition of D2 elements (or loss of D1 elements). Trudgill provided one of the original formulations of salience and Auer et al. (1998) and Kerswill & Williams (2002) developed refinements of Trudgill’s formulation that each represents a significant improvement in the predictive power of the criteria. In this section, I will evaluate the criteria-list approach to salience and discuss the problems and limitations associated with it, ultimately attempting to make the case that the criteria-list approach to defining salience fails to capture the social context in which a dialectal difference is perceived, which, as we will see, is crucial to generating a meaningful, empirically-valid conceptualization of salience.

The most influential and commonly cited of the studies employing a criteria-list definition of salience is Trudgill (1986). He predicted that a linguistic variable will be salient if it meets the four criteria that he provided in his study and further predicted that the variables that are salient will be those that speakers will accommodate to in an interaction with a speaker of another dialect. Auer et al. (1998) refined the criteria-list definition of salience by splitting apart objective criteria from subjective criteria and found that the subjective criteria were more accurate in predicting which variants of a D1 of German would be lost first by speakers acquiring a D2. However, these subjective criteria were not verified to be accurate in empirical tests with speakers and were assumed to be equally reflective of individual speakers’ perceptions. Kerswill & Williams (2002) considered some of Trudgill’s criteria for salience in their study of sound change resulting from dialect contact and concluded that for the notion of salience to avoid circularity in its definition and to be a useful concept for explaining patterns of accommodation and change it must make reference to extra-linguistic factors, although in their description of their model of salience, it is not clear exactly how.
Although Trudgill’s original formulation of salience has received criticism from other researchers and will receive some here as well, it is important to note that his work has been very important in providing a set of testable predictions regarding the effect of salience in dialect acquisition and change and has spawned research that has refined and improved the original definition. Additionally, Trudgill’s formulation of salience is based on empirical research with Jack Chambers (Chambers & Trudgill 1980) in Norwich, England and Trudgill acknowledges that the factors leading to salience in other communities may be different, somewhat softening the specificity of his four criteria. Furthermore, Trudgill’s criteria list was originally intended to explain the pattern of accommodation within a linguistic community, not between communities. That is, Trudgill states that the pattern of accommodation between speakers of two different regional dialects may be very different from the pattern found between speakers of the same dialect. Nevertheless, he does apply his criteria to situations of cross-dialectal accommodation, namely the adaptation of the speech patterns of speakers of British English to American English as a result of living in the United States. Consequently, although the original formulation of the list of criteria for salience stemmed from the observations of Chambers and Trudgill in within-dialect accommodation in Norwich, we may extend the predictions to situations of cross-dialectal interaction, as Trudgill himself does.

### 3.2.1 Social Context

The main problem with using a criteria list to define the salience of linguistic variables is that there often seem to be cases where the criteria do not predict that a particular variable will be salient, yet qualitative evidence suggests that it is. An example of this is the phenomenon known as Canadian Raising. In Canadian English (as well as in some other dialects of English) the diphthongs /aj/ and /aw/ are raised to [ʌj] and [ʌw], respectively, when they occur before a voiceless consonant. For example, *kite* /kajt/ and *couch* /kawtʃ/ are produced as [kʌjt] and [kʌwtʃ] in Canadian English, and as [katʃ] and [kawtʃ] in most dialects of American English. Canadian Raising is a well-known phenomenon and is frequently included in jokes about Canadian English, where the raising process is exaggerated by furthering the raising to the high vowel /u/, giving a realization of the phrase *out and about* as [ut] and [əbut]. That speakers are able to talk about Canadian Raising explicitly and imitate it suggests that is a salient feature of Canadian English. However, it does not seem to fit Trudgill’s criteria for salience.
Pertaining to the first criterion (stigmatization), it is not clear that this feature is stigmatized exactly, since it is not associated with rural or uneducated speech, but rather is associated with an entire dialect; it is a feature of Canadian English and those who use it are easily identified as Canadian, but no judgements about social class or education can be made. Relating to the second criterion (involvement in ongoing sound change), there have been suggestions that Canadian Raising is undergoing change (Chambers & Hardwick 1986), but recent studies suggest that the process is stable (Sadlier-Brown 2011). Trudgill’s third criterion suggests that variables whose variants are radically phonetically different will be salient. The phonetic difference between the unraised diphthongs [aj]/[aw] and the raised versions [Aj]/[Aw] would probably not be considered radical. A change from a low to a mid vowel, while perceptible, is a comparatively small difference within the relatively rich vocalic inventory of English that contains contrasts between vowels that are based not only on height, but also backness and tenseness. Trudgill’s fourth criterion suggests that variables involved in a phonological contrast will be more salient. Since Canadian Raising is a predictable process giving an allophonic realization of the underlying diphthongs, it is not involved in the maintenance of a phonological contrast, although it is part of an allophonic relationship.

It seems then that Canadian Raising does not meet any of Trudgill’s four criteria for salience, yet the inclusion of this feature in jokes about Canadian English, and the fact that those who use it are quickly identified as Canadian, indicates that Canadian Raising is a salient characteristic of Canadian English. Why doesn’t the criteria list capture the salience of Canadian Raising? What can account for the salience of Canadian Raising beyond Trudgill’s criteria for salience?

Part of the reason that the criteria list is unable to capture the salience of Canadian Raising is because, as noted earlier and by other researchers (Auer et al. 1998), it is not always clear how to interpret Trudgill’s criteria in specific instances. In addition, the criteria list is inherently too inflexible to take into account the social context in which a particular linguistic variable is being perceived. The structural characteristics of Canadian Raising do not predict that it will be salient, but studies have found that speakers of other dialects of English seem to have a strong association between the raised variants and Canadian English. For example, Niedzielski (1999) tested the perception of Canadian Raising by Detroit English speakers in order to determine whether listeners use social information, specifically region of origin, in their
perception of phonological variables. She recorded a female Detroit English speaker reading words containing various vowels, including the diphthong /aw/, which is actually realized as the raised variant [ʌw] before voiceless consonants in both Detroit English and Canadian English, and which was realized as the raised variant by the model Detroit speaker. Participants then listened to the words along with synthesized vowels reflecting the same formants as the recording, a more canonical [aw] or a “super-low” variant in which the nucleus of the [a] portion was lower than the canonical [a]. The participants’ task was to select the synthesized vowel that most closely matched the vowel that they heard in the word spoken by the Detroit English speaker. However, only half of the participants were led to believe that the model speaker was a Detroiter; the other half was led to believe that the model speaker was Canadian. Niedzielski found that in the Canadian condition the participants reported that the vowel in the words containing /aw/ most closely matched the synthesized vowel with raised formants 60% of the time, whereas in the Detroit condition they only reported the raised variant 11% of the time. This result was found even though the vowel that was presented to the participants was identical in both conditions and in both conditions was produced by a fellow Detroiter. This finding indicates that Detroiters have a strong association between Canadian Raising and Canadian English, even though the raised variants are also present in their own speech.

This is the cause of the salience of Canadian Raising: the strong association between raising and Canadian English. It is the social meaning of Canadian Raising as a marker of Canadian English that makes it salient. If we consider only the structural characteristics of a linguistic variable, such as whether it is involved in a phonological contrast or in an ongoing sound change, without considering the social context in which the variable is perceived, we may overlook variables that are actually highly salient to listeners, but only in a particular context rather than across the board. Both Detroiters and Canadians have Canadian Raising in their speech, but only when Detroiters believe they are hearing a Canadian does the raising become relevant; indeed, it seems almost not to be perceived otherwise.

Incorporating the idea that salience must be related to social factors, Kerswill & Williams (2002) conclude that for salience to be a predictive notion it must crucially make reference to extra-linguistic factors such as social relations between the two dialect communities in contact, time-scale, intensity of contact, and the involvement of adults or children in the contact situation. As an example of how reference to extra-linguistic factors changes the
definition, consider Trudgill’s first criterion for salience. It states that variants that are stigmatized will be salient, which is likely true, but misses a vital intermediate extra-linguistic, social connection: a variant is stigmatized because it is imbued with negative social meaning reflecting prejudices or stereotypes about a particular social group, frequently low socio-economic classes, or uneducated or rural groups. A stigmatized variant is salient because it reflects this social meaning that speakers of a language understand. It is the social meaning that provides the salience. Kerswill & Williams (2002) note that Trudgill’s *stigmatization* criterion for salience is a sign of salience but not an explanation for it. Kerswill and Williams “see the social psychological property of ‘salience’, which may be attached to a feature by language users, as being linked to internal and extra-linguistic factors…” (Kerswill & Williams 2002: 105). This conclusion is significant because it conceptualizes salience as a social psychological property, rather than as a perceptual or cognitive one. This distinction is crucial because it implies a shift from salience as conceived of as what is noticeable in an acoustic-phonetic sense to what is noticeable in a socially-meaningful sense.

However, even taking into account the conceptual advancement of Kerswill & Williams (2002), the greatest difficulty that the criteria-list approach to defining salience faces is that it is not flexible enough to adjust to the context in which dialectal differences are being perceived. If salience is a social psychological property of linguistic variables, then salience could be relevant for one particular social dimension, while not being relevant for another. For example, as we will see in §3.3, a linguistic variable may be socially salient as a marker of ethnicity, but not as a marker of region of origin. Studies have shown that speakers use particular linguistic variables to signal various social characteristics such as gender, regional dialect, ethnicity, and sexual orientation (Klatt & Klatt 1990; Bachorowski & Owren 1999; Torbert 2004, 2010; Heffernan 2007; Munson 2007; Munson & Babel 2007; Brown 2011). The use of these variables becomes socially meaningful when listeners use them to make judgements about the social characteristics of speakers. Without taking into account social meaning, we may miss the salience of a particular difference if we only follow a list of criteria. For a list of criteria to be sensitive to social context it would need to be indexed for each context in which a variable is perceived, making it completely stipulated for each combination of dialect, linguistic variable, and social context, thereby nullifying the advantage of having a list to begin with.
If we take Kerswill and Williams’ approach that salience is a social psychological property of linguistic variables, then a linguistic variable is salient to the extent that it contributes a particular social meaning, or to the extent that it is used by listeners to make judgements about the social characteristics of speakers. What will be socially meaningful depends on both the presence of particular sociophonetic variation in a speaker’s production and the extent to which a listener is sensitive to that variation. Niedzielski (1999) showed that how a linguistic variable is produced by speakers might not always sync with how it is perceived by listeners: the Detroiter associated raised diphthongs with Canadians, even though the same variants are found in their own speech. Similarly, Laferriere (1979) found that Jewish English speakers in Boston tended to associate the non-standard realization of the orthographic sequence <or> as [ɒ] with the Irish despite the fact that Italians in Boston were found to use [ɔ] more frequently than the Irish in Boston. Hay, Nolan & Drager (2006) found that New Zealand English listeners perceived vowels differently if their answer sheets were labelled with New Zealander or Australian. In this case the actual productions of these two types of speakers are different, unlike in Niedzielski’s study, but it shows that language users do make use of social information, such as regional dialect, when perceiving speech.

3.2.2 Individual Variation in Perception

Another weakness of the criteria list approach to defining salience is that it is not sensitive to individual differences in the perception of linguistic variables. As we have discussed in the preceding sections, Auer et al. (1998) make a distinction between objective and subjective criteria for salience that Trudgill does not explicitly address. The objective criteria are primarily things that linguists know and can describe about a dialect and its variables and, as the criteria are objective, they are presumed to be fixed and independent of any context. In contrast, the subjective criteria are said to reflect the perceptions of language users more directly, although in Auer et al.’s study the subjective criteria were not measured experimentally. Since language users (rather than linguists) are the ones accommodating to other speakers and changing their own production of speech to take on the characteristics of a second dialect, it is reasonable to assume that the characteristics of a particular dialectal variant that make it salient to language users must be known to the language users on some level. That is, what makes a variant salient must be based on the subjective perceptions of language users as well as potentially conscious or subconscious knowledge of language patterns concerning a D2 variant. Language patterns
that speakers have no knowledge of at any level surely cannot be expected to affect how these
speakers will adapt in response to a change in the ambient dialect. The best example of this
problem is the objective criterion for salience of *areal distribution* proposed by Hinskens (1996)
and included as a potential factor in Auer et al. (1998). *Areal distribution* refers to how
geographically widespread a particular dialectal variant is, with the prediction that the more
widespread a variant is, the more salient it will be. It is possible that some language users would
have knowledge about the areal distribution of a dialectal variant, as studies in perceptual
dialectology have suggested (e.g. Preston 1993); but, it must certainly be true that many
speakers would have little or no knowledge about where the geographic boundaries of linguistic
variables would be. Auer et al. (1998) comment on this as well, saying that areal distribution “is
intended to be objective, although the question remains whether dialect speakers’ knowledge of
the areal distribution of certain variables is identical with the dialectologist’s objective data”
(Auer et al. 1998: 167). While linguists can determine what the areal distribution of a dialectal
variant is, it is unlikely that language users would always be able to do so, or that they would do
so in the same way.

Auer and colleagues found that the subjective criteria were more effective than the
objective criteria in predicting which elements of a German D1 would be lost first during
acquisition of a D2. As the subjective criteria would more accurately represent language users’
perceptions of dialectal variants than the objective criteria (by definition), it is perhaps not
surprising that the subjective criteria would perform better in explaining the pattern of dialect
accommodation. Making this distinction between objective and subjective criteria and the
finding that switching from objective to subjective criteria more accurately predicts D2
acquisition represent positive steps towards improving the predictive and explanatory power of a
criteria list for salience. However, that the subjective criteria were not actually defined via
perceptual testing of language users, as mentioned above, denotes a lingering weakness of the
criteria-list approach. Without testing speakers’ perceptions of linguistic variables explicitly
there are two major assumptions made that could be incorrect. First, we are assuming that our
expectation of how speakers perceive a given variable is accurate, even though as noted by
dialects, “surely we are not infallible in these guesses, and without soliciting judgments from the
speakers themselves--another practice fraught with problems--or from speakers outside their
community, who knows exactly what one is or is not aware of, and to what degree?” (Torbert
Second, we are assuming that all speakers perceive the variable in the same way. By definition, the objective criteria are expected to apply equally to all speakers, but the subjective criteria should not necessarily be expected to apply this way. Different groups, such as speakers of different dialects, may find certain variables salient whereas others would not (Hickey 2000). In addition, interspeaker variation in perception that depends on characteristics of the individual also exists (Fox 1983; Makashay 2003), indicating that there may also be patterns that could not be predicted without testing the perceptions of the individual.

3.2.3 Salience as a Gradient Concept

A final limitation of the criteria-list definition of salience is that it is unclear whether this approach allows for salience to be gradient or whether it is limited to a dichotomous attribute: a variant is either salient, or not salient. In the studies discussed earlier in this chapter, linguistic variables which meet the criteria are predicted to be salient, while those that do not meet the criteria are not. Does this suggest that a variable meeting all the criteria is more salient than a variable only meeting some of the criteria? Does meeting each criterion result in an equal degree of increase in salience per criterion or are some criteria weighted more heavily in predicting salience than others? These questions are not addressed in any of the studies incorporating the list-style definition of salience, but some of the criticism of Trudgill’s criteria suggests that in Trudgill’s conception of salience the notion is dichotomous, not gradient. As discussed in §3.1.3, Hinskens (1996) and Kerswill & Williams (2002) consider the idea given in Trudgill (1986) that more salient dialectal variants will be accommodated to, but that variants that have ‘extra-strong salience’ will not be accommodated to, and suggest that such a conception is circular since “salience is sometimes used as an explanation for accommodation…and sometimes to explain why accommodation does not take place” (Hinskens 1996: 11, emphasis in original). If salience is binary – a variant is salient or it is not salient – then Kerswill and Williams’ and Hinskens’ arguments against Trudgill’s notion of ‘extra-strong salience’ are valid, since such a third level of salience is not possible in a binary relation; however, if salience is gradient, with linguistic variables varying in perceptual salience along a continuum, then the notion of extra-strong salience does not result in a circular definition. We might imagine that the relationship between phonetic accommodation and perceptual salience could be linear until a particular threshold of salience is reached, at which point the relationship changes. Trudgill does not explicitly state whether his conception of salience is binary or gradient, but if it is intended
to be gradient, the formulation of salience via a list of criteria does not lend itself easily to reflecting the gradience. In contrast, empirical measures of perceptual salience, specific to a particular social context, capture the gradient nature of salience in a straightforward way, via accuracy rates or ratings along a continuum.

### 3.3 Salience as a Measurable Property of Linguistic Variables

The studies that were discussed in the preceding sections all focused on defining salience via a list of criteria, but none proposed a method of measuring the salience of linguistic variables at the level of the individual language user. Other studies have made use of experimental methods from sociolinguistics and phonetics to attempt to measure the salience of linguistic variables in perception experiments. These studies typically take the approach that the salience of a linguistic variable can be measured by the extent to which the presence of the variable contributes a particular social meaning to listeners. For example, some studies tested listeners’ abilities to accurately identify speaker ethnicity based on the presence of particular features (Graff, Labov & Harris 1986; Thomas & Reaser 2004; Torbert 2004, 2010). Others looked at identifying regional dialect or foreign accents (Boughton 2006; Brunner 2010; Clopper & Pisoni 2004; Fridland, Bartlett & Kreuz 2004; Torbert 2004, 2010). These studies assume that if the presence of a particular feature results in greater accuracy than the presence of another feature in identifying some social characteristic of the speaker who used the feature then the first is more salient than the second within the particular context of identification. This approach provides a concrete step in making the shift discussed previously from salience as what is noticeable in an acoustic-phonetic sense to salience as what is noticeable in a socially-meaningful sense.

Among the first to explicitly test the salience of individual linguistic variables, Graff, Labov & Harris (1986) made use of experimental methods to consider the relative salience of two phonetic variables, /aw/ and /o/, in the identification of ethnicity (Black or White speakers of American English). Both of the variables had been found to have different variants depending on the ethnicity of the speakers who used them in that they tended to be produced as more front by White speakers and farther back by Black speakers. The authors synthesized the vowels to vary between front and back realizations and embedded them in two sentences produced by White speakers, Black speakers, and two speakers who were able to convincingly imitate the other ethnicity’s speech. Listener judges heard each utterance and rated it on a scale of 1 to 7,
where 1 was “sounds very Black” and 7 was “sounds very White” and also decided if the speaker was actually Black or White. They found that the listeners were 3 times more likely to rate a fronted /aw/ variant as “more White” than a more back variant, suggesting that the variants of /aw/ are salient for speakers in identifying ethnicity. In contrast, front and back /o/ produced no significant difference in rating of Whiteness or Blackness, indicating that /aw/ is a more salient variable than /o/ for identifying ethnicity. Graff, Labov & Harris’ study is significant in that it was among the first to empirically test and quantify the salience of linguistic variables and to do so within the context in which language users would perceive the variables: identification of ethnicity. This result does not mean that the difference between front and back /aw/ is inherently more salient than the difference between front and back /o/, but rather that the difference is more socially salient as an identifier of ethnicity for speakers who have experience with the differences. If speakers of other dialects of English where no such marker of ethnicity is found were asked to perform the same experiment using the front and back variants of /aw/ and /o/ the results would likely show no difference in rating as Black or White between the two linguistic variables. The outcome of Graff, Labov & Harris (1986) highlights that the salience of linguistic variables can be measured, but it must be considered within a specific, socially-meaningful context.

Thomas & Reaser (2004) also investigated the identification of ethnicity using similar methodology. They explored whether listeners made use of vowels expected to be diagnostic of ethnicity as well as difference in realization of subject pronouns to identify ethnicity. To create the stimuli, two short speech samples were extracted from longer sociolinguistic interviews. One of the samples contained a vowel that was expected to be salient for ethnicity identification and the other sample did not contain such a vowel, but did contain a subject pronoun. Speech samples were taken from 18 Hyde County, North Carolina speakers: 12 African American speakers and 6 White speakers. The speech samples were presented in random order to listener judges who made decisions about the ethnicity as African American or White of each speaker. The results showed that both the diagnostic vowels and the presence of a subject pronoun were significant predictors of accuracy in identifying the ethnicity of the speaker, suggesting that both vowels and pronouns are salient in this context. The presence of the subject pronoun was also found to contribute more to identification accuracy than the vowels, indicating that the pronouns are more salient than the vowels for ethnicity identification.
Extending this methodology to the identification of regional dialect, Fridland, Bartlett & Kreuz (2004) applied a similar methodology in their study of which aspects of vowel quality are most salient in listeners’ identification of the regional dialect of model speakers, pointing out that “In light of recurrent evidence of the importance of ideology and social affiliation in the spread of sound change, it is important to establish the perceptual dimension of regional dialect differences in terms of what they signal for local speakers” (Fridland et al. 2004: 3). Their study tested whether speakers of Southern US English from Memphis, TN used vowel formants as a cue to the regional dialect of a model speaker (as Southern or Northern) for the high front vowels /i/ and /ɪ/, the mid front vowels /e/ and /ɛ/, the high back vowels /u/ and /ʊ/, and the mid back vowel /ow/. The participants heard two repetitions of a word that included vowels that had been synthesized to reflect either Northern or Southern pronunciation and chose which of the two repetitions sounded more Southern. In addition, the participants rated the similarity of the two repetitions on a numerical scale. The participants’ responses were calculated as %-correct (accuracy). The results showed that the accuracy rate for the front mid vowel /e/ was highest at 84% and the lowest was /ɪ/ at 39% with the other vowels falling in between these two extremes. Fridland and colleagues conclude that /e/ is the most salient vowel for identifying regional dialect between Northern and Southern US English. The variation in accuracy rates suggests that salience of these variables is a gradient concept.

Other studies have incorporated similar methodology to measure the social salience of linguistic variables in different social contexts. For example, Torbert (2004) investigated the perception of three variables of Southern United States English in order to determine how salient each was for the identification of ethnicity and regional dialect. The first variable, /o/, as we have already seen, varies between front and back realizations, with White speakers more likely to produce the fronted variants and Black speakers more likely to produce the back variants. Experiment participants heard speech samples of one or two words containing /o/, with no other potential markers of ethnicity or region, and made a judgment about the model speaker’s ethnicity (White or African American) and region of origin on a scale of 1-5, where 1 was “least Southern” and 5 “most Southern”. Torbert found that White model speakers who produced fronted /o/ were accurately identified by the participants 82% of the time, whereas the African American speaker who fronted /o/ (just one speaker) was correctly identified only 15.8% of the time. Interestingly, lack of fronting (realization of /o/ as back, primarily found in the production of African American speakers) was less salient as a marker for ethnicity with
White model speakers producing back /o/ correctly identified as White 50% of the time and African American model speakers producing back /o/ correctly identified 63.2% of the time. Torbert concluded that the variants of /o/ are salient for identifying ethnicity, but not for identifying region since model speakers who produced front or back /o/ were not classified as being more or less Southern from each other.

Torbert’s study also tested the perception of weak gliding of /ai/. Weak-glided /ai/ is a stereotype of Southern speech and so was expected to be salient for the identification of region, but not for ethnicity since both Black and White speakers in the South have been shown to produce both the glided and weak-glided variant. As expected, weak gliding of /ai/ was found to be socially salient in the identification of region with model speakers producing the weak glide identified with a mean “Southerness” rating of 3.96, which was statistically significantly higher (more Southern) than the gliders rating of 3.29. In contrast, the presence or absence of a weak glide did not affect the identification of ethnicity of the model speakers. These results indicate that weak gliding of /ai/ is salient for identification of region, but not ethnicity.

Torbert’s findings provide crucial evidence that individual linguistic variables can differ in perceptual salience, that their salience can be measured and assigned a specific value, and that salience depends vitally on the social meaning attributed to the difference in variants. What was important for identification of ethnicity (frontness or backness of /o/) was not important for the identification of region (weak gliding of /ai/), and vice versa. In order to know to what extent listeners use particular variables in their assessment of the social characteristics of speakers, we must test their perceptions directly. The salience of linguistic variables depends on the social context in which they are perceived. We cannot use a catch-all diagnostic for salience.

### 3.4 Salience and Phonetic Accommodation

Previous studies considering phonetic accommodation have suggested that the perceptual salience of the various differences between two dialects might affect how speakers of the two dialects will accommodate to each other, although these suggestions are not always in agreement. For example, Evans et al. (in prep) found that speakers of a Northern English dialect of British English diverged away from their Standard Southern British English conversation partners on the most perceptually salient differences between the two dialects. Babel (2010) found that New Zealand English (NZE) speakers converged less (but did not diverge) on the
most salient differences between NZE and Australian English. Trudgill (1986) predicts that speakers of one dialect acquiring a second will converge towards the most salient differences first and most as compared to the less salient differences.

This thesis is concerned with exploring how speakers of two different dialects of Spanish accommodate towards each other on six phonological or phonetic differences between the two dialects. Specifically, I aim to uncover how variation in perceptual salience of these six differences affects the direction and degree of accommodation in conversation. Following the insight of Kerswill & Williams (2002) that salience is a social psychological property of linguistic variables, and incorporating experimental methods employed by Fridland et al. (2004) and other researchers, I define perceptual salience as the extent to which each of the variants of the six dialectal differences between Buenos Aires Spanish and Madrid Spanish contributes to the correct identification of regional dialect by speakers of the two dialects. That is, a dialectal difference will be considered more salient for identifying regional dialect if the listener judges respond more accurately to perceptual trials involving that dialectal difference than they do for trials involving another dialectal difference.

One of the six differences under investigation is the realization of orthographic <c> (before <i> or <e>) and <z>. In Madrid Spanish (MS) these graphemes are realized as the voiceless interdental fricative /θ/, while in Buenos Aires Spanish (BAS) they are realized as the voiceless denti-alveolar fricative /s/. These two sounds are both voiceless fricatives, but they differ in various articulatory and acoustic characteristics, with /s/ being a sibilant and /θ/ not being a sibilant. Whether the difference between /s/ and /θ/ might be considered inherently salient or not is unclear, but the MS variant within this dialectal difference does reflect a well-known characteristic of the dialects of Spanish spoken in Northern and Central Spain (Peninsular Spanish), imbuing the difference between BAS /s/ and MS /θ/ with social meaning. This social meaning stems from the fact that /θ/ in the minds of Spanish speakers is a strong marker of Peninsular Spanish (Piñeros 2009). This association of /θ/ with Peninsular Spanish gives the difference between the realization of orthographic <c> (before <i> or <e>) and <z> social meaning as a marker of regional dialect. The stronger the association, the more meaningful it will be in the minds of Spanish speakers when they hear the use of /θ/.
This social meaning reflects a type of derived salience, separate from that resulting from a syllable becoming more prominent when word-initial or stressed, for example. Since this dialectal difference is frequently cited in the literature as being a strong marker of MS, I predict that listener judges who are native speakers of Spanish will respond with high accuracy to trials involving this dialectal difference. The accuracy rate with which they respond will be taken as a reflection of the perceptual salience of that dialectal difference in the social context of identifying regional dialect. It is expected that the six dialectal differences will vary in how salient they are for identifying regional dialect. As eloquently put by Torbert (2004) “since human beings tend to know themselves and their in-group well and tend to paint other groups with few, broad strokes, a small number of highly salient features can come to characterize a "discrete" variety in the popular consciousness, while other features may escape notice or have relatively less overt social marking attached to them” (Torbert 2004: 1).
Chapter 4
The Dialects: Buenos Aires Spanish and Madrid Spanish

The two dialects of Spanish that were chosen to be included in the experiment to investigate phonetic accommodation in cross-dialectal conversation in Spanish are Buenos Aires Spanish and Madrid Spanish. Buenos Aires Spanish (BAS) is spoken in the city of Buenos Aires, which is the capital of Argentina with a population of just under 13 million (Instituto Nacional de Estadística y Censos 2010). The city itself is an autonomous region within the country. Madrid Spanish is spoken in the city of Madrid, which is the capital and the largest city of Spain with a population of around 3.3 million. Statistics from the Instituto Nacional de Estadística (Nacional Institute of Statistics) of Spain tell us that 83% of the population of Madrid is from Spain and 17% is made up of immigrants. The most common source of immigration to Madrid is Ecuador making up 15% of the immigrant population. The remainder of the immigrant population includes people from many different countries, including Argentina, which comprises almost 2% of the immigrant population. Madrid Spanish (MS) has much in common with the Spanish spoken in the rest of northern and central Spain (Peninsular Spanish), including a large overlap in consonantal inventories and identical phonemic vowel inventories (Piñeros 2009: 135-143). For the purposes of the present study it might not have been necessary to isolate MS as a dialect separate from the rest of Northern-Central Peninsular Spanish, but since the experimentation all took place in Madrid it was straightforward to restrict the participants to those born and raised in Madrid, thereby avoiding any potential differences between speakers within the Peninsular Spanish group relating to dialect.

This particular pair of dialects makes a good test case for investigating cross-dialectal phonetic convergence for three main reasons. The first is that there are many differences in the sound systems of BAS and MS, of which the articulatory and acoustic properties have been studied previously (Borzone de Manrique & Massone 1981; Harris & Kaisse 1999; Hualde & Prieto 2002; Aguilar 2003, 2005; Martínez Celdrán, Fernández Planas & Carrera Sabaté 2003; Face & Alvord 2004; Hualde 2005; Chang 2008; Hualde, Simonet & Torreira 2008; Piñeros 2009; Colantoni & Kochetov 2011), providing the necessary knowledge to examine shifts in speech patterns by speakers of these dialects. The second reason is that the various differences in sound systems between BAS and MS are predicted to vary in perceptual salience, which is necessary in order to consider the role of salience in the pattern of phonetic accommodation.
The third reason is that migration from Argentina to Madrid is common, as noted above (Instituto Nacional de Estadística 2007), so speakers of BAS and MS, living in the same city, would regularly have the opportunity to come into contact with each other and potentially affect each other’s speech. Testing phonetic accommodation in conversation between speakers of BAS and MS then mirrors a process which could take place in the real world during contact between speakers of these two dialects in Madrid.

The phonemic vocalic inventories of BAS and MS are the same, made up of the 5 vowel system shown in Figure 1.

\[
\begin{align*}
/i/ & \quad /u/ \\
/e/ & \quad /o/ \\
/a/ &
\end{align*}
\]

Figure 1: The phonemic vowel system of Spanish

The phonemic consonantal inventory of BAS is given in Table 3 and the inventory of MS is given in Figure 4. The inventories of the two dialects have much in common, as well as a couple of notable differences.

<table>
<thead>
<tr>
<th>Table 3: Consonantal phonemic inventory of Buenos Aires Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bilabial</strong></td>
</tr>
<tr>
<td>Plosive</td>
</tr>
<tr>
<td>Nasal</td>
</tr>
<tr>
<td>Trill</td>
</tr>
<tr>
<td>Tap</td>
</tr>
<tr>
<td>Fricative</td>
</tr>
<tr>
<td>Affricate</td>
</tr>
<tr>
<td>Lateral</td>
</tr>
</tbody>
</table>

The primary differences between the inventories are that MS has the voiceless interdental fricative /θ/, while BAS does not, and BAS has the voiceless (or voiced, for some speakers) postalveolar fricative /ʃ/, while MS does not. Where BAS speakers produce /ʃ/, MS speakers produce the palatal consonant /j/. This symbol /j/ technically corresponds to the voiced palatal
fricative, but in MS the realization of this sound varies along a range of degrees of constriction including the glide [j], fricative [ʝ], stop [ɟ], and affricate [ɟʝ], but the voiced palatal weak fricative or approximant [ʝ] is the most common realization in Standard Castilian (Hualde 2005: 165).

Table 4: Consonantal phonemic inventory of Madrid Spanish

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Interdental</th>
<th>Denti-Alveolar</th>
<th>Alveolar</th>
<th>Postalveolar</th>
<th>Palatal</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>p b</td>
<td>t d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>k g</td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td></td>
<td></td>
<td></td>
<td>n</td>
<td>j</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tap</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>f θ</td>
<td></td>
<td></td>
<td></td>
<td>s</td>
<td>j x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affricate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>l</td>
<td></td>
<td></td>
<td>(ʎ)2</td>
</tr>
</tbody>
</table>

The voiced stops /b d g/ participate in an alternation with the continuant approximant allophones [β δ γ], where the stops appear in utterance-initial position and after homorganic consonants and the approximants appear elsewhere (Harris 1969; Hwu 1994; Amastae 1995; Romero 1995; Widdison 1997; Carreira 1998; Face 2002; González 2002). In discussions of Spanish phonology, the palatal glide [j] and the labiovelar glide [w] are not usually included in the consonant inventory of Spanish. Instead, the glides are usually argued to be allophones of the high vowels /i/ and /u/, respectively (Roca 1991; Harris & Kaisse 1999), which surface when unstressed and adjacent to another vowel, such as in fiesta /fiesta/ [ˈfjes.ta] ‘party’ or bueno /bueno/ [ˈbwe.no] ‘good’. In Martínez Celdrán, Fernández Planas & Carrera Sabaté (2003)’s discussion of Castilian Spanish, the derived palatal glide found in diphthongs with preceding consonants is treated as a separate phoneme from the voiced palatal affricate /ʝ/ found in onset position (with no preceding consonant), such as in yate /ʝate/ ‘yacht’, which is realized as the affricate in utterance-initial position and after laterals or nasals, while in all other contexts it surfaces as the approximant [ʝ]. In the current study, the symbol /j/ will be used, following Hualde (2005), to represent the palatal consonant whose most common realization is a palatal

2 The palatal lateral is still part of the consonant inventory for those who speak a lleísta dialect of Spanish, one that contrasts /ʎ/ with the palatal consonant /ʝ/, such as in areas of Northern Spain and the Andean highlands (Whitley 2002). In the central highlands of Ecuador the contrast between /ʎ/ and /ʝ/ is maintained, but the former is realized as a palatoalveolar fricative [ʒ] instead of as a palatal lateral (Lipski 1994). Yeísta dialects, those that have neutralized this contrast to /ʝ/, are much more commonly spoken, particularly among younger speakers (Hualde 2005: 56).
approximant in MS. Although the symbol [j] is a fricative, since the most common realization is an approximant in MS, throughout the thesis I will refer to this consonant as an approximant.

Spanish syllables consist of an obligatory nucleus and an optional onset and coda. The nucleus can be made up of a singleton vowel as in y [i] ‘and’ or a sequence of a glide+vowel (rising diphthong), as in pie [pjε] ‘foot’ or vowel+glide (falling diphthong), as in hay [aj] ‘there is/are’, or less commonly a triphthong as in buey [bwej] ‘ox’. The onset and coda can consist of at most two consonants and their distributions are quite limited (Hualde 2005: 71). For example, clusters of /s/ plus another consonant are not permitted in onset position. In word-medial clusters, the /s/ would be syllabified as the coda of the preceding syllable, while the second consonant would be syllabified as the onset of the following syllable. There are similar types of rules guiding the syllabification of sequences of two (or more) vowels. Under certain circumstances a sequence of two vowels is permitted to remain in hiatus, while in others the sequence is syllabified as a diphthong. Variation in how these vocalic sequences are syllabified will constitute two of the differences between BAS and MS that are included in the present study and will be discussed in §4.

Spanish stress is phonologically contrastive, giving rise to minimal pairs such as hablo [‘a.βlo] ‘I speak’ and habló [a.’βlo] ‘he/she spoke’, which differ only in the placement of stress. Stress must fall on one of the final three syllables of the word. In words ending in a consonant the unmarked pattern is oxytonic (stress falling on the final syllable), as in tensión [ten.'sjon] ‘tension, tautness’, while in words ending in a vowel the unmarked pattern is paroxytonic (stress falling on the second to last syllable), as in persona [per.'so.na] ‘person’ (Hualde 2005: 221).

Beyond the distinctions in consonant inventory, there are many differences between BAS and MS spanning multiple domains of speech, including lexical items, pronominal inventory and usage, verb formation, intonation contours, patterns of vocalic sequences, and acoustic and articulatory properties of shared segments. This thesis focuses on the differences between the dialects in terms of segmental inventories, treatment of vocalic sequences, and the acoustic characteristics of two phones that both dialects have. Six dialectal differences in particular are included and the extent to which these differences undergo alignment in a conversation involving a speaker from each dialect area is investigated. These six differences are listed in Table 5 and will be discussed in individual subsections.
Table 5: List of dialectal differences between Madrid Spanish and Buenos Aires Spanish

<table>
<thead>
<tr>
<th>#</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The realization of orthographic &lt;c&gt; (before &lt;e&gt;, &lt;i&gt;) and &lt;z&gt; as [s] in BAS, but as [θ] in MS</td>
</tr>
<tr>
<td>2</td>
<td>The realization of orthographic &lt;y&gt; and &lt;ll&gt; as [ʃ] (or [ʒ]) in BAS, but [ʝ] in MS</td>
</tr>
<tr>
<td>3</td>
<td>The laminal realization of /s/ in BAS versus the apical realization of /s/ in MS</td>
</tr>
<tr>
<td>4</td>
<td>The velar realization of /x/ in BAS versus the uvular realization of /x/ in MS</td>
</tr>
<tr>
<td>5</td>
<td>The presence of diphthongization of mid-vowel sequences in BAS, but not in MS</td>
</tr>
<tr>
<td>6</td>
<td>The presence of exceptional hiatus in MS, but not in BAS</td>
</tr>
</tbody>
</table>

4.1 Dialectal Difference #1 (BAS /s/ versus MS /θ/)

Difference #1 reflects the realization of orthographic <c> (before <e> or <i>) and <z>. In MS, these graphemes are produced as the voiceless interdental fricative /θ/ (Martínez-Celdrán et al. 2003; Hualde 2005: 20; Piñeros 2009: 136), but this phoneme is not found in BAS. Where MS speakers produce /θ/, BAS speakers produce the voiceless denti-alveolar fricative /s/, such as in the word *pozo* ‘well’, realized as [ˈpo.θo] in MS, and as [ˈpo.so] in BAS. Since in MS there is a phonological contrast between /θ/ and /s/ and in BAS this contrast is merged to /s/, this dialectal difference reflects one of the four factors given by Trudgill (1986: 11) (discussed in §3.1) for a linguistic variable to be highly salient and thus, susceptible to accommodation. In addition, the phoneme /θ/ is often mentioned as being strongly associated with the Spanish of Spain (e.g. Piñeros 2009: 136), and as such, is a marker of many dialects of Spanish spoken there (including MS), suggesting that Difference #1 will be salient.

Fricatives are typically characterized in the literature by spectral moments, frication duration, overall amplitude, or amplitude relative to a following vowel (Borzone de Manrique & Massone 1981; Dart 1993; Hedrick & Ohde 1993; Jongman, Wayland & Wong 2000; Gordon, Barthmaier & Sands 2002; Kochetov & Lobanova 2007; Li, Edwards & Beckman 2009; Li, Munson, Edwards, Yoneyama & Hall 2010). The four spectral moments are centre of gravity, standard deviation, skewness, and kurtosis. Centre of gravity refers to the location of the average energy of the frication, standard deviation is a measure of the range of energy about the mean, skewness is a measure of how evenly distributed the energy is above and below the mean, and kurtosis is a measure of how much the shape of the distribution differs from a Gaussian.

---

3 The voiceless interdental fricative also occurs as an allophone of /d/ in word-final position in MS, such as in *Madrid* /maˈðɾiθ/ (González 2002); however, only consonants in onset position are considered in the present study.
shape with respect to the well-formedness of the spectral peaks (Jongman et al. 2000; Heffernan 2004).

Centre of gravity (COG) has been shown to be an important acoustic cue for contrasts in place of articulation of coronal fricatives (Jongman et al. 2000). In general, the value of COG is negatively correlated with the size of the resonating cavity in front of the oral constriction, such that more anterior fricatives, with a smaller front resonating cavity, have higher COG than those with a more posterior articulation (Forrest, Weismer, Milenkovic & Dougall 1988; Li et al. 2010). Although previous studies giving COG values for Spanish /s/ and /θ/ are not available, Jongman et al. (2000) report that English /θ/ has a higher COG than English /s/. Using COG to differentiate between /s/ and /θ/ is not completely straightforward, however, since the majority of the acoustic studies of fricatives have focused on isolating the cues used to discriminate between fricatives within the sibilants or between fricatives within the non-sibilants, but not on discrimination across these groups. Relative amplitude of the fricative may prove a more useful cue to this particular contrast than COG. Hedrick and Ohde (1993) found that the amplitude of a fricative relative to the amplitude of the following vowel was a robust cue to perception of place of articulation in the English contrasts /s/-/θ/ and /s/-/ʃ/. Focusing on the /s/-/θ/ contrast, in that study, stimuli were synthesized in terms of relative amplitude to range from /s/ to /θ/ and participants listened to the stimuli and identified which sound they heard. The authors found that there were more /s/ responses when the relative amplitude was higher and there were more /θ/ responses when the relative amplitude was lower. A similar result was found in Strevens (1960). Following these results, we expect that MS /θ/ will have lower relative amplitude than BAS /s/ and we use this measure to characterize Dialectal Difference #1. Details on exactly how relative amplitude was measured can be found in §5.3.1.1.

4.2 Dialectal Difference #2 (BAS /ʃ/ versus MS /ʝ/)

Difference #2 concerns another difference between the segmental inventories of the two dialects. In MS, orthographic <ll> and <y> are produced as the palatal approximant /ʝ/. In contrast, these graphemes are realized as either the voiceless alveopalatal fricative /ʃ/ or the voiced counterpart [ʒ] in BAS, with the voiceless realization most frequent among younger

---

4 Again, as noted in the first section of this chapter, although the symbol [ʝ] is technically a fricative, since the most common realization of this segment in MS is the glide [j], I will refer to this sound as an approximant or glide.
speakers (Hualde 2005: 31; Chang 2008). For example, the word *llego* ‘I arrive’ is pronounced ['ʝe.yo] in MS, but as either ['ʝe.yo] or ['ʐe.yo] in BAS.

As discussed by Martinez-Celdrán et al. (2003), MS /j/ is expected to be realized as the approximant [j] in all contexts except utterance-initial position and after nasal or lateral consonants, where the stronger affricate [ʝ] is expected. While these two allophones are those given by Martinez-Celdran et al. (2003), Hualde (2005: 165) points out that, in fact, MS /j/ can vary along a range of degrees of constriction including the glide [j], fricative [ʝ], stop [ɟ], and affricate [ɟʝ], but that the voiced palatal weak fricative or approximant [j] is the most common realization in Standard Castilian, although Hualde also notes that in colloquial MS, the stronger realizations of /j/ may also be found in contexts where the weaker versions might be expected, such as in intervocalic position. In general, though, we expect the stronger variants of MS /j/ in strong positions, such as word initially or after laterals or nasals, while the weaker variant (the approximant [j]) is expected elsewhere.

This variation in the realization of MS /j/ may reflect a difference in consonantal strength. Similarly, Difference #2 (BAS /ʃ/ versus MS /ʝ/) involves a contrast\(^5\) between two segments that might also be thought of as differing in consonantal strength, with MS /ʝ/ being voiced and more vowel-like (at least in most contexts), and therefore weaker, as compared to BAS /ʃ/ (Escure 1977; Balakrishnan, Freyman, Chiang, Nerbonne & Shea 1996; Lavoie 2001). Studies considering the range of possible surface realizations of /g/ in Argentine Spanish have found that there is a relationship between consonantal strength and relative intensity of the consonant, where relative intensity is defined as the difference between the intensity of the consonant and the intensity of the following vowel. MacLeod (2010) found that for BAS /g/, the stop allophone [g] had a lower relative intensity than the phonologically weaker approximant realization [ɣ], which in turn had lower relative intensity than the weakest variant of /ɡ/, the labiovelar glide [w]. Following this pattern of lower relative intensity for phonologically stronger consonants, we would expect that BAS /ʃ/ should have a lower relative intensity than MS /ʝ/, at least when the latter is produced as the approximant allophone. As such, relative intensity is used here to distinguish between BAS /ʃ/ and MS /ʝ/. The minimum intensity of the approximant (MS /ʝ/) or fricative (BAS /ʃ/) and the maximum intensity of the following vowel

---

\(^5\) The term ‘contrast’ here is not used in the sense of a phonological contrastive relationship, but rather in its more general sense.
were measured in order to calculate the relative intensity of the consonant, which is defined as the difference between the intensity of the consonant and the following vowel (Balakrishnan et al. 1996; MacLeod 2010, 2012; Mazzaro 2011).

Harris & Kaisse (1999) and Barrancos (2008) note that the fricatives [ʃ] and [ʒ] are strongly associated with the Spanish of Argentina and allow BAS to be easily recognized, suggesting that these fricatives are perceptually salient for dialect identification. In the contexts where MS /ʝ/ is realized as a variant that is stronger than the approximant (such as a fricative or affricate), the difference between MS /ʝ/ and BAS /ʃ/ may be less distinct.

4.3 Dialectal Difference #3 (BAS Laminal [s] versus MS Apical [s̺])

Both Differences #3 and #4 concern differences in the articulatory and acoustic properties of two fricatives that are common to BAS and MS. Difference #3 reflects the difference in realization of the voiceless denti-alveolar fricative /s/. In Northern and Central Spain, encompassing Madrid, the most common realization of /s/ is the apico-alveolar [s̺], whereas the predominant realization in the rest of the Spanish-speaking world, including in Buenos Aires, is the lamino-alveolar [s] (Hualde 2005: 47; Piñeros 2009: 136). For example, the word sala ‘classroom’ is realized as ['s̺a.la] in MS, but as ['sa.la] in BAS.

There are three main differences in articulation between these two sounds: first, during production of the MS apical [s̺], the apex of the tongue points upward, whereas for the BAS laminal [s] it points downward; second, the point of constriction of apical [s̺] is somewhat farther back than that of the laminal [s]; third, in production of the apical [s̺] the narrow passageway that is formed between the tongue and the alveolar ridge where the air flows is rounded, whereas during production of the laminal [s] this passageway is flatter (Piñeros 2009: 272). Perceptually, the difference between the apical and laminal /s/ is that the apical [s̺], due to its more posterior place of articulation, sounds more similar to the voiceless alveopalatal fricative /ʃ/ of English, although that sound is articulated even farther back than the MS apical [s].

The salience of the difference between BAS laminal [s] and MS apical [s̺] is likely lower than that between BAS /s/ and MS /θ/ since the difference between BAS and MS /s/ is within
the category of anterior sibilants, while BAS /s/ and MS /θ/ span the categories of sibilants and non-sibilants. We expect that between-category differences will be easier to discriminate than within-category differences, as predicted by the Perceptual Assimilation Model (PAM: Best 1995). However, the apical realization of MS /s/ was one of the most commonly imitated features of Peninsular Spanish in Reiter (2004)’s study of Spanish accent imitation, which according to Trudgill’s criteria for salience should render the difference salient.

Another factor that may cause Difference #3 to be less salient than Difference #1 is that Difference #1 is reflected in the orthography, while Difference #3 is not. Trudgill (1986: 11) notes that variables that do not sync with the orthography are often stigmatized, which is one of his criteria for a variable to be considered salient. It may be the case that the more a particular variant syncing with the orthography the less salient it may be. As discussed, /θ/ is used to realize orthographic <c> (before <e> or <i>) and <z> in MS. While there is nothing inherent about these graphemes that suggests they should be pronounced as the voiceless interdental fricative, rather than any other segment, the distribution is completely predictable once learned. In MS, these graphemes are produced as /θ/ and <s> is produced as /s/. BAS speakers could acquire this pattern simply by learning which graphemes require a /θ/ realization. This situation of having a predictable, but not inherently logical, association with the orthography creates an intermediate level between the logical /t/-voicing example given by Chambers (1992), which was discussed in §2.1.2.2, and a situation where a contrast is completely neutralized in the orthography and is only lexically specified. For example, in Canadian English the pair of words bead [bid] and head [hed] have the same orthographic sequence of letters <ea> to represent two different vowels. There is nothing about the orthography that signals whether the vowel is /i/ or /ɛ/. This contrast is only lexically defined, it does not sync logically, and it is not a distribution that could be learned from the orthography. That bead has /i/ as its nucleus and that head has /ɛ/ as its nucleus must be acquired on an item by item basis. This would be a situation where the orthography cannot be used at all to learn the distribution, potentially making such a difference more difficult to acquire. Difference #3 (BAS laminal [s] versus MS apical [ʂ]) is not signalled in the orthography in the way that Difference #1 (BAS /s/ versus MS /θ/) is. Even though the association with MS /θ/ with orthographic <c> (before <e> or <i>) and <z> is not inherently logical, its predictable distribution may contribute to making Difference #1 more salient than Difference #3.
As noted above, centre of gravity is an important acoustic cue for contrasts in place of articulation of coronal fricatives (Jongman et al. 2000), being negatively correlated with the size of the resonating cavity in front of the oral constriction. Since BAS /s/ has a more anterior realization than MS /s/, it is expected to have a higher COG than the MS counterpart. In addition to the denti-alveolar fricative /s/, BAS also has the voiceless postalveolar fricative /ʃ/ (or /ʒ/ depending on the age of the speaker (Hualde 2005: 31; Chang 2008)). Since this fricative is articulated farther back than the alveolar ridge, it is expected that this fricative would have a lower COG than either the BAS /s/ or the MS /s/.

4.4 Dialectal Difference #4 (BAS Velar [x] versus MS Uvular [χ])

Difference #4 concerns the difference in articulation of the voiceless velar fricative /x/ between the two dialects. In Madrid, the tendency is to realize this fricative with a more posterior post-velar or even uvular place of articulation and with more turbulent frication than the velar fricative produced in Buenos Aires (Hualde 2005: 154; Piñeros 2009: 136). /x/ in Buenos Aires and many other areas of Central and South America is a weaker and less turbulent realization of this segment as compared to that in Madrid. For example, the word jarra ‘pitcher’ is realized as [ʼxa.ɾa] in BAS and as [ʼχa.ɾa] in MS. Like Difference #3, Difference #4 is likely not as salient as Differences #1 and #2 which are cross-category differences, while Differences #3 and #4 are more within-category differences in articulation. In addition, the dialect-specific realization of /x/ was not found to be commonly imitated in Reiter’s (2004) study of Spanish dialect imitation.

As noted in Ladefoged & Maddieson (1996: 166) and Werker & Tees (2002), there is very little research published on the acoustic differences between velars and uvulars, but Gordon et al. (2002) examined various acoustic measures such as duration, centre of gravity, and the value of the first and second formants at the onset of a vowel following a fricative to characterize voiceless fricative contrasts in seven diverse languages. One of the languages they considered, Montana Salish, has a phonemic contrast between uvular /χ/, labialized velar /xʷ/, and labialized uvular /χʷ/. They found that these fricatives did not differ significantly in duration or centre of gravity (overall intensity was not measured), but did differ significantly in the value of the second formant (F2) at the onset of the following vowel: the farther back the constriction in the oral cavity, the lower F2. This result aligns with patterns of F2 in vowel production as well. Vowels that are articulated farther back in the oral cavity have lower F2 than vowels.
articulated farther front (Lieberman 1977; Stevens 1998; Johnson 2003; Ladefoged 2005: 42). If the primary difference between BAS [x] and MS [χ] is that the MS variant is articulated somewhat farther back, then we would expect that the value of F2 at the onset of the vowel following the fricative will be lower for MS speakers than for BAS speakers.

4.5 A Note on Spanish Vocalic Sequences

Differences #5 and #6 concern the treatment of sequences of two vowels. Before we proceed with discussing these two differences, it will be useful to define the terms that will be used extensively in the next two sections and to review some background research on vocalic sequences.

In Spanish, a sequence of two vowels in which one is a high vowel can be syllabified as a diphthong (monosyllabic glide+vowel sequence), such as [ja] or [we], or as a hiatus (heterosyllabic sequence of two vowels) such as [i.a] or [u.e] (Harris 1969; Hualde 1991, 1997, 1999, 2004; Roca 1997; Aguilar 1999; Hualde & Prieto 2002; etc.). Some analyses of Spanish propose that the glides [j] and [w] are allophones of the high vowels /i/ and /u/, respectively, and that the glide variants surface when the high vowels are adjacent to a stressed vowel (Roca 1991; Harris & Kaisse 1999). This approach gives a general rule of syllabification where a sequence of two vowels in which one is high is realized as a diphthong if the high vowel is unstressed, such as in *piara/*piara/ [ˈpja.ɾa] ‘herd’ or *cuero/*kuero/ [ˈkwe.ɾo] ‘leather’, and as a hiatus when the high vowel is stressed, such as in *día/*dia/ [ˈdi.a] ‘day’ or *bambúes/*bambues/ [bam.ˈbu.es] ‘bamboos’ (Borzone de Manrique 1979; Quilis 1993; Hualde & Prieto 2002).

While the phonological definition of diphthongs and hiatus is clear, phonetic definitions of these categories are more difficult to establish (Senturia 1998). Leichte & Peterson (1961) propose that a diphthong is “a vocalic syllable nucleus containing two target positions”. A hiatus, following this definition, would also contain two target positions, but would consist of two syllable nuclei. Within the vocalic sequence we find three elements: the onset steady state, the transition from initial to final vowel (or vocoid), and the offset steady state (Borzone de Manrique 1979). Speaker intuitions regarding the syllabification of vocalic sequences (as diphthong or hiatus) have been shown to be fairly robust (at least for Peninsular Spanish: Hualde & Prieto 2002; Face & Alvord 2004), and several experimental studies have attempted to determine what the acoustic correlates of these intuitions are. Building on the observation that
hiatus are typically longer than diphthongs (Borzone de Manrique 1979; Quilis 1981; Aguilar 1999), Hualde & Prieto (2002) compared the syllabification intuitions and productions of six native speakers of Madrid Spanish to determine to what extent the durational difference agrees with the intuitions. They found that in five of the six participants the agreement was high. The durations of the diphthongs were shorter, on average, than the durations of the hiatus, for each speaker. Face & Alvord (2004) found that speakers of Barcelona Spanish were able to distinguish between diphthongs and hiatus with considerable accuracy on the basis of duration.

4.6 Dialectal Difference #5 (Diphthongization of Mid Vowel Sequences in BAS, but not in MS)

When a sequence of two vowels does not contain a high vowel (so a sequence of two non-high vowels), the general rule of syllabification (Quilis 1993) predicts that the sequence will always be realized as a hiatus, and this is generally the case in MS. Dialectal Difference #5 concerns the difference in treatment of vocalic sequences of two non-high vowels in MS and BAS. In contrast to MS, in BAS and many other Latin American dialects, sequences in which the first vowel is a mid vowel (and the second is mid or low) variably undergo diphthongization such that the mid vowel is raised and realized as a glide (Jenkins 1999; Hualde 2005: 23; Garrido 2007; Hualde et al. 2008). For example, a word such as pelear /pelear/ ‘to fight’ can be realized as [pe.le.'ar] or as [pe.'ljaɾ]. In Madrid, while there is variation between these two realizations, particularly in colloquial speech, there is a tendency to preserve the pronunciation containing the hiatus, whereas in Latin American dialects the diphthong realization is common even in careful speech (Navarro Tomás 1970: 68; Hualde 2005: 23). Part of the reason for this asymmetry between the dialects could be that the raising of mid vowels to a glide is heavily stigmatized in Peninsular Spanish (Hualde et al. 2008). According to Alonso (1930), diphthongization of mid vowel sequences has been much more accepted in Latin American varieties than in Peninsular Spanish due to a greater prestige afforded to the oral use of the language, or how speakers genuinely use the language, in Latin American countries than in Spain.

Garrido (2007) investigated diphthongization of mid vowel sequences in dialects of differing prestige of Colombian Spanish and found that the process was affected by dialect, stress, and task formality. In that study, 17 monolingual Spanish speakers from Colombia (8 speakers from the Andean region and 9 from the Caribbean region) read a list of target words and narrated a story from pictures. The stimuli included in both tasks contained either the
orthographic sequence <io> (predicted to be realized as [jo] by all speakers) or <eo> (predicted to be realized as [e.o] more often by the Andean speakers and as [jo], showing diphthongization of the sequence, by the Caribbean speakers). The stimuli were controlled for their position relative to stress and included vocalic sequences in pre-tonic, tonic, or post-tonic position. In the tonic position stress fell on the second element of the vocalic sequence. Supporting previous results that diphthongs tend to be shorter than hiatus (Aguilar 1999; Hualde & Prieto 2002; Face & Alvord 2004; MacLeod 2007, 2012) and incorporating the logic that the vowel /e/ raised to [j] would have a lower F1 than [e], Garrido found that the more innovative Caribbean speakers showed a greater tendency to diphthongize the sequence /eo/ to [jo] than the Andean speakers. Furthermore, the dialectal difference was more apparent in the more formal reading task than in the story narration task. She concludes that the coastal speakers use diphthongization as a way to mark their identity as speakers of that dialect, while the Andean speakers avoided diphthongization to signal their membership in the higher prestige group.

Following the results of Garrido (2007) and Hualde et al. (2008), we predict that diphthongization of mid vowel sequences in the current study will be reflected in a lowering of the value of the first formant in the first vowel in the sequence (/e/) and a shortening in the duration of the entire vocalic sequence. In addition, this change may also be reflected in a raising of F2, since [e] is somewhat farther back than [i] (Stevens 1998: 277).

4.7 Dialectal Difference #6 (Potential Exceptional Hiatus in MS, but not in BAS)

The general rule of syllabification discussed above predicts that a sequence of two vowels containing a high vowel that is unstressed will always be realized as a diphthong (Quilis 1993). Many studies have investigated the so-called ‘exceptional hiatus’ found in some varieties of Spanish, including Peninsular Spanish (subsuming Madrid Spanish), in which a hiatus is produced even when the high vowel is not stressed, as in cliente /kliente/ [kli.ˈen.te] ‘client’. Since the high vowel is not stressed, following Quilis’ rule of syllabification, the expected realization of the vocalic sequence is the diphthong [je] instead of the hiatus [i.e], giving [ˈkljen.te]. Exceptional hiatus is not found in BAS; the possibility of exceptional hiatus in MS but not in BAS is Dialectal Difference #6.
Exceptional hiatus seem to be found in a particular set of lexical items and Hualde (1997, 1999) discusses some of the morphological and phonological patterns evident in that set. Exceptional hiatus are often found when there is a morphologically related word in which the high vowel is stressed. For example, the word *viable* /bi.a.ble/ [bi.'a.ble] ‘viable’ contains an exceptional hiatus and is related to the word *via* /bia/ [bi.a] ‘way’ in which the high vowel /i/ is stressed. In addition, vocalic sequences that are separated by a morphological boundary, such as between a stem and an affix, are often realized as exceptional hiatus. For example, the sequence /ie/ in *bienio* /bienio/ [bi.'e.njo] ‘two-year period’ is split by a boundary between the prefix /bi/ and the stem and has been identified as containing an exceptional hiatus. In contrast, the word *viento* /biento/ [bjen.to] ‘wind’ contains the same underlying vowel sequence in a very similar phonological context, and yet the sequence is syllabified as a diphthong (Face & Alvord 2004). Also, when a vocalic sequence is in word-initial position or when it follows a word-initial consonant, hiatus is favoured, particularly when that initial consonant is /r/, such as in *riendo* /riendo/ [ri.'en.do] ‘laughing’ as compared to *siendo* /siendo/ [sjen.do] ‘being’ (Hualde 1999).

Even when these factors are taken into account there are still exceptional hiatus that seem to be truly exceptional, evading all attempts to explain their distribution. The difference between MS and BAS is that exceptional hiatus are much more common in MS than in BAS (Hualde 2005: 82). A word like *riendo* ‘laughing’ is predicted to be produced as [ri.'en.do] by speakers of MS, but as ['rjen.do] by speakers of BAS. The acoustic correlate of this difference is duration, according to the many results investigating this contrast in the literature (Aguilar 1999; Hualde & Prieto 2002; Face & Alvord 2004; Macleod 2007, 2012). The vocalic sequence in the MS pronunciation, if produced as an exceptional hiatus, should be longer than that in BAS, which is realized as a diphthong.
Chapter 5
Methodology

To investigate the pattern of phonetic accommodation along the six differences between the dialects of BAS and MS discussed in §4 and evaluate the role of perceptual salience in phonetic accommodation, I conducted an experiment in which pairs of Spanish speakers participated in a series of tasks. This chapter details the participants who took part in the study (§5.1), the experimental procedure and stimuli (§5.2), and the analysis of the data (§5.3).

5.1 Participants

In total, 11 pairs of participants took part in the experiment. In each pair there was one speaker from Madrid and one from Buenos Aires. The participants were recruited from various sources including posters advertising the experiment and posted on university campuses in Madrid, via online social networking sites, such as Couchsurfing (http://www.couchsurfing.com), and via my existing contacts in Madrid.

All participants were 18 years of age or older, with university education completed or in progress, did not report hearing problems, and were from either Madrid or Buenos Aires. The participants from Buenos Aires varied in the length of time they had lived in Madrid. A chart detailing the age and gender of the participants is shown in Table 6 below. For the BAS speakers, the length of time they had lived in Madrid and whether or not they had plans to stay are also given. The mean amount of time living in Madrid across all BAS participants was just under 1.5 years (17.6 months) and the median was 6 months. Seven of the 11 BAS participants indicated that they were planning to stay in Madrid; the others were students who planned to return to Argentina when they finished school.
Table 6: Gender and ages of both participants per pair, and length of time in Madrid for the BAS speakers.

<table>
<thead>
<tr>
<th>Pair</th>
<th>Gender</th>
<th>MS speaker Age</th>
<th>Age</th>
<th>BAS speaker</th>
<th>Time lived in Madrid</th>
<th>Plans to stay in Madrid?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>39</td>
<td>35</td>
<td>2 months</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>26</td>
<td>27</td>
<td>6 months</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>23</td>
<td>32</td>
<td>1 year</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>40</td>
<td>27</td>
<td>1 month</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>31</td>
<td>30</td>
<td>6 months</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>26</td>
<td>25</td>
<td>2 months</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>25</td>
<td>52</td>
<td>8 years</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>37</td>
<td>29</td>
<td>2.5 years</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>31</td>
<td>26</td>
<td>2.5 years</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>36</td>
<td>34</td>
<td>3 months</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>F</td>
<td>32</td>
<td>19</td>
<td>6 months</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

Each pair consisted of either two men or two women in order to avoid asymmetrical effects of gender (e.g. Namy et al. 2002). The experiment took a total of almost 2 hours to complete and each participant was compensated €20 for his or her time.

5.2 Outline of Tasks

The general methodology included the use of a map task (Anderson et al. 1991), which was flanked by the reading of a word list, in order to compare baseline production with post-conversation production to identify any changes. This method of comparing baseline and post-exposure production has been used by various researchers including Pardo (2006), Evans & Alshangiti (2010), and Nielsen (2011). The steps of the experiment are illustrated in Figure 2 below and explanations of each task follow.

Figure 2: Outline of tasks performed by each participant in the experiment
All of the experimentation took place in Madrid, Spain in the quiet, multi-purpose space of a rental apartment, which included two adjacent rooms. The general procedure was explained to each pair of participants at the beginning of the experiment. In addition, they read an information letter regarding the experiment and signed an informed consent form. The participants’ speech was recorded using two Fostex FR-2 digital recorders and two Audio-Technica AT831b lavalier microphones. During the word list, perception task, and repetition task, the participants were seated behind a laptop, each in a separate room. To listen to the stimuli in the perception and repetition tasks, the participants used Sennheiser HD280 headphones. During the conversation portion of the experiment the participants were seated on opposite sides of the same table and were both recorded onto one of the Fostex FR-2 recorders, using two microphones on two separate channels.

5.2.1 Step 1: Baseline (Pre-Conversation) Production Task

The first step of the experiment was to establish a baseline of pronunciation for each participant. This was accomplished through a simple word reading task. The Spanish words used as stimuli were embedded in the carrier phrase *Digo __________ para ti* “I say __________ for you” and were presented in written form to the participants in a PowerPoint presentation in which the slides advanced automatically every 3 seconds. The participants’ task was simply to read the sentences aloud. The stimuli included in the word list were 167 Spanish tokens, presented in a different random order for each participant, and were selected in order to provide samples of the production of the segments involved in the six dialectal differences under investigation. Explanation of the criteria that the stimuli met, exceptions, and examples are discussed below.

5.2.1.1 Stimuli: Fricatives

The stimuli chosen to investigate the dialectal differences that pertain to fricatives (Difference #1: BAS /s/ vs. MS /θ/, Difference #2: BAS /ʃ/ vs. MS /ʝ/, Difference #3: BAS laminal [s] vs. MS apical [s], and Difference #4: BAS velar [χ] vs. MS uvular [χ]) all contained the target segments in onset position and were controlled for the following characteristics:

1. *Position of the target segment:* always in onset position, but either word-initial or word-medial
2. *Location of stress:* target segment was either the onset of the stressed syllable or not
3. *Quality of the following vowel:* stimuli were balanced by following vowel /a,e,i,o,u/
For each of Differences #1 to #4 we expect 20 stimuli since there are 2 positions (word-initial or word-medial) x 2 stress patterns (stressed or unstressed) x 5 following vowels. In order to evaluate whether the different orthographic forms of Differences #1 and #2 affected their realization, I have attempted to include 20 stimuli per orthographic representation of the target segment. In some cases there are not enough real Spanish words to make up the desired 20 per group. The stimuli are broken down below and any shortfalls are noted in Table 7.

Table 7: Breakdown of stimuli included in experiment for Differences #1 - #4

<table>
<thead>
<tr>
<th>Diff.</th>
<th>Description</th>
<th>Orthographic Form</th>
<th># of Stimuli</th>
<th>Example</th>
<th>Gloss</th>
<th>MS</th>
<th>BAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BAS /s/ v. MS /θ/</td>
<td>&lt;z&gt;, &lt;ce&gt;, &lt;ci&gt;</td>
<td>14, 16</td>
<td>zona</td>
<td>zone, area</td>
<td>['θo.na]</td>
<td>['so.na]</td>
</tr>
<tr>
<td>2</td>
<td>BAS /ʃ/ v. MS /ʝ/</td>
<td>&lt;y&gt;, &lt;ll&gt;</td>
<td>17, 17</td>
<td>yoga</td>
<td>yoga</td>
<td>['ʝo.ya]</td>
<td>['ʃo.ya]</td>
</tr>
<tr>
<td>3</td>
<td>BAS [s] v. MS [s]</td>
<td>&lt;s&gt;^6</td>
<td>20</td>
<td>sala</td>
<td>classroom</td>
<td>['sala]</td>
<td>['sela]</td>
</tr>
<tr>
<td>4</td>
<td>BAS [x] v. MS [χ]</td>
<td>&lt;j&gt;, &lt;gi&gt;, &lt;ge&gt;</td>
<td>20</td>
<td>jarra</td>
<td>pitcher</td>
<td>['χa.ra]</td>
<td>['xa.ra]</td>
</tr>
</tbody>
</table>

A total of 20 stimuli were not included for the <z>, <y>, and <ll> target segments because stimuli containing <yi>, <lli> or word-medial <ze> are extremely infrequent in Spanish (Davies 2002) and would likely have elicited an unnatural production. Only 8 stimuli for orthographic <ce> or <ci> are expected since there are 2 positions (initial versus medial), 2 stress positions, and only 2 following vowels, however, I have included 16 stimuli (2 per cell) in order to provide a larger sample than only 8 tokens.

5.2.1.2 Stimuli: Vocalic Sequences

The 39 stimuli chosen to investigate diphthongization of mid vowels in sequence all contained one of the following VV-sequences: /io/, /eo/, /ia/, /ea/. The /io/ and /ia/ tokens were expected to be realized as diphthongs by both dialects and are included as controls to compare with the production of /eo/ and /ea/ tokens. The stimuli were controlled for the following characteristics:

1. *Position of the sequence*: initial (always with a preceding onset consonant) or medial
2. *Location of sequence with respect to stress*: pretonic, tonic or posttonic

^6 Of course, in BAS, orthographic <z>, and <c> when /e/ or /i/ follow are also realized as /s/. In this experiment, I separated BAS /s/ into two categories, those where MS speakers would also produce /s/ (pertaining to Difference #3) and those where MS speakers would produce /θ/ (pertaining to Difference #1).
Since a pretonic initial sequence is not possible, the total number of stimuli expected is 20. In order to ensure a large enough sample of productions to evaluate diphthongization, 2 stimuli per cell were included, except for the initial tonic /io/ sequence, which is very rare in Spanish (Davies 2002). For this sequence, only one stimulus was found: dioses ‘gods’ /di.o.ses/.

The 16 stimuli investigating production of diphthongs and hiatus were made up of pairs of phonetically-similar words identified in the literature as containing a diphthong and an exceptional hiatus in Peninsular Spanish (thus expected to be a diphthong in BAS) (Hualde 1991; Hualde & Prieto 2002; Face & Alvord 2004). Examples of the stimuli containing vocalic sequences are provided in Table 8 and the complete stimulus set is given in Appendix A.

Table 8: Breakdown of stimuli included in experiment for Differences #5 - #6

<table>
<thead>
<tr>
<th>Diff.</th>
<th>Description</th>
<th>Sequence</th>
<th>N</th>
<th>Example</th>
<th>Gloss</th>
<th>MS</th>
<th>BAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Diphthongization of midV sequences in BAS, not in MS</td>
<td>/io/</td>
<td>9</td>
<td>violín</td>
<td>violin</td>
<td>[bjo.'lin]</td>
<td>[bjo.'lin]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>/eo/</td>
<td>10</td>
<td>leopardo</td>
<td>leopard</td>
<td>[le.o.'par.do]</td>
<td>[ljo.'par.do]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>/ia/</td>
<td>10</td>
<td>copia</td>
<td>copy</td>
<td>['ko.pja]</td>
<td>['ko.pja]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>/ea/</td>
<td>10</td>
<td>linea</td>
<td>line</td>
<td>['li.ne.a]</td>
<td>['li.nja]</td>
</tr>
<tr>
<td>6</td>
<td>Exceptional hiatus in MS, not in BAS</td>
<td>Expected diphthong</td>
<td>12</td>
<td>diente</td>
<td>tooth</td>
<td>['djen.te]</td>
<td>['djen.te]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Potential hiatus</td>
<td>12</td>
<td>cliente</td>
<td>client</td>
<td>[kli.'en.te]</td>
<td>['kljen.te]</td>
</tr>
</tbody>
</table>

5.2.2 Step 2: Perception Task

As the primary focus of this thesis is determining to what extent perceptual salience affects phonetic accommodation, it was necessary to establish a metric by which perceptual salience could be quantified. As discussed in §3, I incorporated experimental methods employed by Fridland et al. (2004) and other researchers to define perceptual salience as the extent to which each of the variants of the six dialectal differences contributes to the correct identification of regional dialect by speakers of the two dialects. That is, a dialectal difference will be considered more salient for identifying regional dialect if the listener judges respond more accurately to perceptual trials involving that dialectal difference than they do for trials involving another dialectal difference.

In a typical AX discrimination task, participants hear two stimuli and decide whether they are the “same” or “different”. Using an AX task to establish whether or not the participants
could hear the difference between the dialectal differences would not be sufficient in this case since the goal of this experiment is the question of whether the dialectal differences differed in magnitude. If one dialectal difference is greater than another, then after hearing trials for that dialectal difference and perceiving a large difference, the participants may expect a large difference for trials involving other dialectal differences, causing them to judge the less salient dialectal differences as the “same”, even if they could actually perceive some acoustic difference. To avoid this, it may have been possible to have used an AXB task, but preliminary testing suggested that the memory load was too great and participants expressed difficulty completing the task, a difficulty that other researchers have faced with the AXB task as well (Evans & Alshangiti 2010). Instead, perception of the dialectal differences was tested via a combined discrimination and identification task in which the participants determined which of two recordings sounded most like their native dialect. The model recordings included 8 native speakers of Spanish (4 BAS: 2 males and 2 females, and 4 MS: 2 males and 2 females). All but one speaker (one male MS speaker) were recruited and recorded in Toronto. The other was recorded in Madrid. The stimuli recorded are discussed in §5.2.2.1.

The perception experiment was created and run in Praat (Boersma & Weenink 2009). During the task the participants saw the orthographic form of a stimulus on the computer screen (the characteristics of the stimuli are discussed in the following section) and heard two pronunciations of the stimulus. Their task was to decide which of the two repetitions reflected the pronunciation norms of their native dialect area. There were four possible combinations of pronunciation of the two repetitions:

1. Both repetitions were produced by speakers of MS
2. Both repetitions were produced by speakers of BAS
3. The first repetition was produced by a speaker of MS and the second by a speaker of BAS
4. The first repetition was produced by a speaker of BAS and the second by a speaker of MS

In the cases where both repetitions were produced by speakers of the same dialect, the two repetitions were always produced by two different speakers of the same dialect; no voice was repeated twice in any trial.

The participants saw the orthographic form of the stimulus on the screen, heard the two repetitions and then had five possible responses:
1. that only the first repetition sounded like their native dialect  
2. that only the second repetition sounded like their native dialect  
3. that both repetitions sounded like their native dialect  
4. that neither repetition sounded like their native dialect  
5. that they do not know whether either repetition sounded like their native dialect

For example, in the sample trial shown in Figure 3 below, the participants saw the stimulus *azá* and, after a 2-second delay, heard two pronunciations of it, separated by 200ms of silence. This short inter-stimulus interval (ISI) was chosen in order to allow comparison at an acoustic-phonetic level rather than at a phonemic level, following the results of Werker & Logan (1985) and Brannen (2002) that suggested that an ISI of less than 250ms would allow phonetic processing rather than phonological processing. In the current study phonetic processing was desired since comparing some of the dialectal differences (such as Differences #3 and #4) at a phonological level could disallow use of the finer-grained phonetic differences that distinguish the MS and BAS variants. In the example trial in Figure 3, if the first pronunciation was [a.'θa] and the second one [a.'sa] the participants could potentially make use of the difference to determine that the first pronunciation is more typical of Madrid Spanish than the second, since the phoneme /θ/ is not found in BAS. The participants were able to click a button at the bottom of the screen to listen to the repetitions up to 3 times and in the event that they accidentally clicked an unintended response button they could click a button at the right-hand side of the screen to go back to the previous trial and correct the error.

Figure 3: *Sample trial of the perception task to determine perceptual salience*†

†This sample trial shows the screen as the MS participants would have seen it. For the BAS participants the screen would have read “Primero de Buenos Aires”, “Segundo de Buenos Aires”, etc.
The purpose of the perception task was to determine how strongly speakers of MS and BAS associate certain production patterns with their own dialect. Following the logic and findings of Fridland et al. (2004), the prediction was that the participants would perform more accurately on the trials involving the more salient dialectal differences than on those containing the less salient differences.

The measure used to quantify perceptual salience was the percentage of the trials involving each dialectal difference that each participant responded correctly to: %-Correct.

5.2.2.1 Perception Task Stimuli

The stimuli included sequences of sounds that captured the realization of the target segments. For the dialectal differences involving fricatives (Differences 1-4) the stimuli included three different forms: CV (stressed initial), 'V.CV (unstressed medial), and V.'CV (stressed medial). The vowel included in these stimuli was /a/. The stimuli for the dialectal differences involving vocalic sequences (Differences 5-6) contained a sequence of vowels that could potentially be produced as a diphthong or a hiatus, depending on the dialect. In some cases these stimuli formed genuine Spanish words such as in the case of Dialectal Difference #2 (BAS /ʃ/ vs. MS /ʝ/), where the target segment in unstressed medial position gives the word [a.'ʃa] for BAS and [a.'ʝa] for MS, which corresponds to the word allá ‘there’. In other cases the stimuli formed nonce words such as in the case of Dialectal Difference #1 (BAS /s/ vs. MS /θ/) where the target segment in initial position generates [ˈsa] for BAS and [ˈθa] for MS, which are not genuine Spanish words. In two cases the stimuli are real words in one dialect, but not in another. Both of these cases are found within Difference #1. When the target segments are in unstressed medial position we get [ˈa.sa] for BAS, which corresponds to the real word asa ‘handle’, but [ˈa.θa] for MS, which is not a real word. Similarly, when the target segments are in stressed medial position, the stimuli created are [a.'sa] for BAS, which is the realization of the word asá ‘roast.imperative’, but [a.'θa] in MS, which is not a real word. However, in both of these cases the orthographic forms of the stimuli (which are visible to the participants in the perception task) are different from how the real words would be spelled since the stimuli would be spelled aza and azá, rather than asa and asá.

Although it would be better to have purely nonce stimuli for the perception task, it is extremely difficult to generate natural, short sequences of sounds for all of the 6 dialectal
differences where none of the sequences correspond to genuine Spanish words. Although consistency in whether the stimuli were real Spanish words or not would be preferred, a comparison of the accuracy rates of the various trials found that there was virtually no difference in accuracy between the trials in which the stimuli were real words and those in which they were nonce words for either dialect. On the trials involving real words, the BAS speakers responded correctly to 20% and the MS speakers responded correctly to 22% of the tokens. On the trials involving nonce words, the BAS speakers responded correctly to 23% and MS speakers responded correctly to 27%. A logistic mixed effects regression model (which will be discussed in more detail in §5.3.5) found no significant effect of the authenticity of the word for either dialect, indicating that any effect of the status of the stimuli as real or not is minimal in the current study.

The stimuli are listed below in Table 9 including the MS and BAS pronunciations of each. Recordings were made of 4 MS and 4 BAS speakers (2 males and 2 females of each) reading these nonce words aloud, both embedded in the carrier phrase *Digo _________ para ti* and without a carrier phrase. Those produced without the carrier phrase were used in the stimuli to avoid unnatural sounding stimuli resulting from the target word being excised from a longer production. An individual sound file was made for each token, with mean intensity normalized at 70dB.
Table 9: Breakdown of stimuli included in perception task

<table>
<thead>
<tr>
<th>Difference</th>
<th>(Nonce) Word</th>
<th>MS</th>
<th>BAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>za</td>
<td>[θa]</td>
<td>[sa]</td>
</tr>
<tr>
<td></td>
<td>aza</td>
<td>['a.θa']</td>
<td>['a.sa']</td>
</tr>
<tr>
<td></td>
<td>azá</td>
<td>[a.'θa]</td>
<td>[a.'sa]</td>
</tr>
<tr>
<td>2</td>
<td>ya, lla</td>
<td>[ja]</td>
<td>[a], [ʒa]</td>
</tr>
<tr>
<td></td>
<td>aya, alla</td>
<td>['a.ja']</td>
<td>['a.ʒa], ['a.ʒa']</td>
</tr>
<tr>
<td></td>
<td>ayá, allá</td>
<td>[a.'ja]</td>
<td>[a.'ja], [a.'ʒa]</td>
</tr>
<tr>
<td>3</td>
<td>sa</td>
<td>[ sjə]</td>
<td>[sa]</td>
</tr>
<tr>
<td></td>
<td>asa</td>
<td>['a.ʃə]</td>
<td>['a.sa']</td>
</tr>
<tr>
<td></td>
<td>asá</td>
<td>[a.'ʃə]</td>
<td>[a.'sa]</td>
</tr>
<tr>
<td>4</td>
<td>ja</td>
<td>[ʃə]</td>
<td>[xa]</td>
</tr>
<tr>
<td></td>
<td>aja</td>
<td>['a.ʃə]</td>
<td>['a.xa]</td>
</tr>
<tr>
<td></td>
<td>ajá</td>
<td>[a.'ʃə]</td>
<td>[a.'xa]</td>
</tr>
<tr>
<td>5</td>
<td>óteo</td>
<td>['o.te.o]</td>
<td>['o.tjo]</td>
</tr>
<tr>
<td></td>
<td>atea</td>
<td>['a.te.a]</td>
<td>['a.tja]</td>
</tr>
<tr>
<td></td>
<td>teotó</td>
<td>[te.o.'to]</td>
<td>[tjo.'to]</td>
</tr>
<tr>
<td></td>
<td>teatá</td>
<td>[te.a.'ta]</td>
<td>[tja.'ta]</td>
</tr>
<tr>
<td>6</td>
<td>diá</td>
<td>[di.'a]</td>
<td>[dja]</td>
</tr>
<tr>
<td></td>
<td>riá</td>
<td>[ri.'a]</td>
<td>[rja]</td>
</tr>
<tr>
<td></td>
<td>tiotó</td>
<td>[ti.o.'to]</td>
<td>[tjo.'to]</td>
</tr>
<tr>
<td></td>
<td>tiatá</td>
<td>[ti.a.'ta]</td>
<td>[tja.'ta]</td>
</tr>
</tbody>
</table>

In the perception task to establish perceptual salience, each stimulus was presented in 3 trials: once with two MS realizations, once with two BAS realizations, and once with one MS and one BAS realization (balanced for MS-BAS and BAS-MS orders). This totals 81 trials, which were presented interspersed with 51 control trials, in a different random order for each participant. The 51 controls included three trials each of the 17 control tokens shown in Table 10, once with two presentations by MS speakers, once with two presentations by BAS speakers, and once with one presentation by an MS speaker and one by a BAS speaker, evenly balanced between MS-BAS and BAS-MS orderings. The control trials contained various consonants and vowels, but none of the six dialectal differences or any other known difference between BAS and MS. The purpose of the control trials was to ensure that the participants were not making use of other cues not related to the six dialectal differences in making their judgements about the dialect of the model speaker in the perception task. Like the test stimuli, some of the control stimuli form genuine Spanish words, but as noted above, none contain any of the dialectal differences that we
expect might allow the participants to perceive the stimuli as having been produced by a speaker of BAS or MS. A logistic mixed effects regression model found no significant difference between the accuracy rates of the real control trials and the nonce control trials for either dialect (real: BAS = 19%, MS=22%; nonce: BAS = 23%, MS = 28%).

Table 10: Control stimuli used in perceptual task to determine perceptual salience.

<table>
<thead>
<tr>
<th>Difference</th>
<th>(Nonce) Token</th>
<th>MS</th>
<th>BAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ba</td>
<td>[ba]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>da</td>
<td>[da]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>ga</td>
<td>[ga]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>pa</td>
<td>[pa]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>ta</td>
<td>[ta]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>ka</td>
<td>[ka]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>ma</td>
<td>[ma]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>na</td>
<td>[na]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>la</td>
<td>[la]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>ti</td>
<td>[ti]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>te</td>
<td>[te]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>to</td>
<td>[to]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>tu</td>
<td>[tu]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>di</td>
<td>[di]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>de</td>
<td>[de]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>do</td>
<td>[do]</td>
<td>Same as MS</td>
<td></td>
</tr>
<tr>
<td>du</td>
<td>[du]</td>
<td>Same as MS</td>
<td></td>
</tr>
</tbody>
</table>

The stimuli in the perception task that were intended to reflect MS pronunciation were produced by MS speakers and those intended to reflect BAS pronunciation were produced by BAS speakers, so we expect that these stimuli would actually be reflective of their dialect, but to ensure that there was a difference between the BAS and MS tokens, statistical comparisons of the relevant acoustic measures were conducted. The results showed that each of the dialectal differences differed between the dialects significantly and in the expected direction. The output for the models is given in Appendix D.

5.2.3 Step 3: Repetition Task

If convergence in terms of one of the six dialectal differences between MS and BAS investigated in this thesis is not found, it could be because the speakers find the opposing dialect
variant too difficult to produce, introducing a limitation on accommodation that has nothing to do with how salient the dialectal difference is. In order to verify that the participants were able to produce the six dialectal differences included in this study, the third task that they performed was a repetition task using the rapid shadowing technique (Goldinger 1998). In rapid shadowing participants hear a recording and immediately repeat it with no delay between the offset of the recording and onset of repetition. Goldinger (1998) found that shadowed tokens with no delay were imitated more faithfully than those in which a 3-4 second delay was introduced, even when no instruction to imitate was given. Other studies have found evidence of spontaneous imitation using rapid shadowing as well (Shockley et al. 2004; Mitterer & Ernestus 2008). Shadowing with no delay will give the participants the best possible chance to reproduce the recordings as faithfully as possible. If they are unable to produce the sounds not found in their native dialect during the shadowing task, then convergence along those measures will be very unlikely due to articulatory complexity. To maximize the likelihood of imitation (thereby being able to evaluate the participants’ ability to produce the sounds of the opposing dialect), participants were instructed to repeat the stimuli they heard as similarly as possible, even if that meant using sounds they would not normally use.

The same stimulus set used in the perception task was used in the repetition task. Participants heard 40 test trials (1 of each possible MS and BAS realization of the stimuli given in §5.2.2.18), with 3 repetitions of each trial, along with 34 distractor trials (each of the 17 distracters produced once by an MS speaker and once by a BAS speaker), for a total of 222 trials in random order with no two identical trials appearing consecutively. The repetition experiment was presented via headphones and run through Praat (Boersma & Weenink 2009). The orthographic form of the stimuli was not presented to the participants; they only heard the recording of the stimuli and repeated. A discussion of the results of the repetition task is found in §6.1.8.

5.2.4 Step 4: Conversation

After completing the baseline production task, the perception task, and the repetition task, the pairs of participants engaged in conversation together to provide exposure to the contrasting

---

8 Since the tokens for the BAS realization of Differences #1 and #3 are the same (/s/), the stimuli for the BAS realization of these differences were only included once.
dialect. The conversation component was made up of two parts as detailed in the following sections.

5.2.4.1 Map Task

The first part was a map task (Anderson et al. 1991), which has been used successfully by Pardo (2006) in her investigation of spontaneous convergence in conversation. The map task involves the use of a pair of maps generated by the researcher that include labeled landmarks whose names contain the specific segments or characteristics under investigation. Within the pair of maps, one includes a route drawn between two points that passes by the various landmarks. The other map contains the landmarks, but not the route. The task is for the two participants to communicate with each other such that the person with the map missing the route is able to reproduce the route as similarly to the provided route as possible. The map task is well suited to studying phonetic convergence in conversation since the participants are focused on the task at hand, rather than their own pronunciation patterns, and the custom-naming of the landmarks allows for the inclusion for the specific segments under investigation. Furthermore, the effect of having the conversation is readily evaluated via comparison of the baseline and post-conversation production tasks (in this case, a word list).

One difficulty with the map task is that occasionally the conversation can be somewhat one-sided, with the participant holding the map with the route (the giver) providing instructions and the other (the receiver) drawing the route and listening, but not speaking. In order to facilitate two-way conversation the participants were instructed that when receiving they should repeat the giver’s instructions as they received them, ostensibly to ensure that the instructions were clear with the ultimate goal of making the two routes as similar as possible. Including this type of repetition ensures that both participants are speaking during the task, providing each participant with significant exposure to the other’s pronunciation.

Each pair of participants performed the map task twice, once each as giver and once as receiver, in order to balance asymmetry effects related to role in a conversation (Giles, Coupland & Coupland 1991; Shepard et al. 2001; Pardo 2006; Pardo et al. 2010). Each map took 10-15 minutes to complete. The participants were seated on either side of a table. The maps were fixed to clipboards which the participants held such that it was not possible to see each other’s maps. The speakers were recorded according to the specifications given in §5.2.
5.2.4.2 Guided Conversation

The second part of the conversation component was less structured than the map task and allowed the participants to engage in more natural dialogue that could include longer stretches of speech (thereby providing more exposure to the opposing dialect for each participant), more eye contact, and a chance to develop rapport. The topic suggested by the researcher was differences between the cities of Madrid and Buenos Aires with respect to architecture, culture, nightlife, lifestyle, safety, etc. The BAS participant generally began by describing some of things that he or she had noticed that were different between the two cities and by making recommendations on places to see and things to do in Buenos Aires for the MS participant. After approximately 7-10 minutes the MS participant was asked to take the lead and talk about some of the aspects of life in Madrid that he or she particularly liked and to make suggestions to the BAS participant for things to do in Madrid. Although the role as information provider changed between these two segments, both participants were actively involved in discussing the differences between the two cities in both parts of the conversation.

5.2.5 Step 5: Post-Conversation Word List

In order to determine whether phonetic accommodation had taken place and whether it was affected by the salience of the dialectal difference, the participants performed the same word-list reading task as in the baseline phase immediately after the conversation component was completed. The productions pre- and post-conversation were compared acoustically to determine whether any changes had occurred and whether any such changes reflected convergence towards (or divergence away from) the opposing dialect. This comparison methodology has been used in many studies, including Abrego-Collier et al. (2011), Nielsen (2011), Black (2012), and Mielke, Magloughlin & Nielsen (2012).

5.2.6 Step 6: Questionnaire

The final step of the experiment was the completion of a questionnaire designed to collect the participants’ personal data such as year and place of birth and education level along with their experience with Madrid and Buenos Aires, opinions regarding how the Argentine community is viewed in Madrid, knowledge of the differences between the two dialects, and reactions about how they would feel if their Spanish changed as a result of moving to another dialect area. The questionnaire is provided in Appendix C.
5.3 Data Analysis

This section details how the acoustic analysis was conducted (in §5.3.1), how phonetic accommodation was assessed (in §5.3.2), how the magnitude of the changes made from pre- to post-conversation were standardized across the six dialectal differences (in §5.3.3) and the types of statistical analyses that were used (in §5.3.4).

5.3.1 Acoustic Analysis

For the production data, each sound file containing the pre- or post-conversation word list for each of the participants was manually segmented and analyzed using a script in Praat (Boersma & Weenink 2009) to calculate the relevant measures for the six dialectal differences investigated. These measures are relative intensity (RI) (for Differences #1 and #2), centre of gravity (COG) (for Difference #3), mean value of F2 at the onset of the following vowel (onset F2) (for Difference #4), duration of the vocalic sequence (for Difference #6), and a composite value reflecting both duration of the vocalic sequence and height of the first vocoid\(^9\) (for Difference #5).

5.3.1.1 Relative Intensity

As explained in §4.1 and §4.2, Difference #1 (BAS /s/ vs. MS /θ/) and Difference #2 (BAS /ʃ/ vs. MS /ʝ/) were differentiated using relative intensity (RI). The minimum intensity of the approximant (MS /ʝ/) or fricatives (BAS /ʃ/, BAS /s/ or MS /θ/) and the maximum intensity of the following vowel were measured in order to calculate the relative intensity of the approximant or fricative and the intensity of the following vowel to calculate relative intensity (Balakrishnan et al. 1996; MacLeod 2010, 2012; Mazzaro 2011). The fricatives were manually marked in a Praat textgrid at the onset and offset of aperiodic noise in the waveform.

5.3.1.2 Centre of Gravity

As discussed in §4.3, centre of gravity (COG) is an appropriate measure for distinguishing between the apical and laminal realizations of /s/ (Difference #3). Since BAS /s/ has a more anterior realization than MS /s/, it is also expected to have a higher COG than the MS

\(^9\) F2 was also measured, but as will be discussed in §6.1.2, turned out not to be a significant predictor of the difference between the BAS and MS realizations of mid vowel sequences.
counterpart. COG was calculated from a 30ms Hamming window centered around the midpoint of the frication noise (Kurowski, Hazan & Blumstein 2003). Fricatives were manually marked in a Praat textgrid at the onset and offset of aperiodic noise in the waveform.

5.3.1.3 Mean Onset F2

The value of the second formant (F2) at the onset of a vowel following a fricative has been shown to be negatively correlated with the length of the resonating cavity behind the constriction (Halle and Stevens 1997; Stevens, Li, Lee & Keyser 2004). As such, the mean value of F2 at the onset of the vowel following /x/ was used to distinguish between the velar realization of the BAS [x] and the more posterior articulation of the MS [χ], with the expectation that the BAS [x] would have a higher onset F2 than MS [χ]. Onset F2 was estimated using an linear-predictive coding algorithm implemented in Praat (Boersma & Weenink 2009) for 5 formants that was calculated over a range from 0 to 5000Hz for men and 0 to 5500Hz for women, with a window length of 0.025 (Li et al. 2009; Li et al. 2010).

5.3.1.4 Vocalic Sequence Duration

For Differences #5 (diphthongization of midV sequences in BAS, but not MS) and #6 (potential exceptional hiatus in MS, but not BAS) the difference in syllabification is expected to be reflected in a durational difference since hiatus tend to be longer than diphthongs (Aguilar 1999; Hualde & Prieto 2002; Face & Alvord 2004; MacLeod 2007, 2012). In tokens involving vocalic sequences, the total duration of the sequence was measured. The onset and offset of the vocalic sequences were manually marked in Praat, following cues from both the waveform and the spectrogram. The start of the sequence was marked at the first regular vocal pulse, at the zero crossing, and the end of the sequence was marked at the offset of F2 (Chitoran 2002; Limanni 2008).

Since measures of duration are dependent on speech rate, such data must be normalized before they can be compared across groups or speakers. The durations of the vocalic sequences were normalized to control for differences in speech rate via the following procedure. The mean word duration was calculated for each speaker across all tokens. The durations of the vocalic sequences were then adjusted to bring the mean word duration for each speaker into line with that of the mean word duration of all speakers together. The mean word duration of all the speakers together was 454ms. To normalize the duration measurements for the BAS speaker
from Pair 4, for example, whose mean word duration was 398ms, a scaling factor of \(1 + \frac{(454 - 398)}{454}\) = 1.12 was applied to all of the durations of vocalic sequences taken for this speaker, bringing his mean word duration up to 454ms. This procedure was also used in MacLeod (2012) and allows the duration measurements taken from different speakers to be compared directly. Any difference in duration then will reflect a genuine difference, rather than one that stems from variation in speech rate.

5.3.1.5 Composite Sequence Duration + Vocoid Height

While the literature suggests that the most robust acoustic correlate of the diphthong/hiatus contrast is duration of the sequence, in the case of mid vowel sequences (hiatus) that have been diphthongized, thereby raising the initial [e] to [j], we might also expect to see a change in formant values for that vocoid. In particular, the vocoid should have lower F1 if diphthongization takes place, since F1 is inversely correlated with the height of the tongue during production of a vocoid, as well as a higher F2 (Lieberman 1977: 67; Stevens 1998: 277; Johnson 2003: 112; Ladefoged 2005: 42). As such, the acoustic measure used to characterize the difference between mid vowel sequences realized by BAS and MS speakers would likely more accurately reflect the dialectal difference if both sequence duration and F1 and F2 of the first vocoid were taken into account. The method by which these measures were combined into a single composite measure is described in §6.1.5, since it depends on the actual productions of these sequences by the participants in the study.

Sequence duration for the tokens reflecting Difference #5 was measured in the same way as those for Difference #6, described in the previous section. To measure F1 and F2, the first vocoid in the sequence was manually marked at its onset and offset. Although Aguilar (1999) notes that diphthongs do not always have two clear steady states, the majority of sequences in this study were found to have two steady states, although the first (corresponding to the glide) was often very short. A Praat script created a formant object using the command To Formant (burg), then excised a 30ms portion at the midpoint of the vocoid and mean F1 and F2 were calculated.

5.3.2 Assessing Phonetic Accommodation

Previous studies investigating phonetic accommodation have used various methods to assess the direction and degree of accommodation. The particular method chosen depends on the specific
goals of the study. For example, in Pardo (2006), the main goal was to determine whether speakers would accommodate towards each other during a conversational interaction, without focusing specifically on particular sounds or features of speech. Instead, Pardo was interested in getting a broad measure of convergence, incorporating whichever acoustic characteristics of the signal that listener judges would use to detect convergence. For her purposes, the AXB task was appropriate. In the AXB task, listener judges decide whether a token that a speaker produces late in the conversation is more like the conversational partner’s production than the same token produced by the speaker earlier in the conversation. If listener judges choose the token produced late in the conversation more often than the early one, this indicates that by the end of the conversation the speaker sounded more like his partner. Using this method does not allow us to say on what basis the listener judges make their decisions. Whichever acoustic cues are present in the signal are available for them to use, in some particular configuration and weighting. The AXB method of assessing perceptual salience gives a holistic measure of convergence in which multiple acoustic-phonetic cues are likely integrated. This method has been used in other studies of phonetic imitation as well, including Namy, Nygaard & Sauerteig (2002), Kim, Horton & Bradlow (2011), and Babel, McGuire, Nicholls & Walters (2012).

Because in the current study I am interested in how much a particular acoustic correlate of a dialectal difference has changed from pre- to post-conversation, acoustic analysis was instead conducted. Other studies of phonetic imitation have also used acoustic analysis where a particular acoustic measure was needed. For example, Shockley et al. (2004), Abrego-Collier et al. (2011), and Nielsen (2011) measured VOT and Babel (2012) measured F1 and F2 of vowels. In this study, as discussed in §4, the six dialectal differences between BAS and MS were each defined via a particular acoustic dimension (or combination of dimensions), and these dimensions were measured according to the processes described in §5.3.1, for both the pre-conversation production and the post-conversation production. With these measurements in hand, the next step was to compute the amount that each participant had changed on each of the six dialectal differences from pre- to post-conversation, giving the magnitude of the change, as well as to determine the direction of the change (convergence or divergence) relative to the productions of the opposing dialect.

In terms of the magnitude of the change, this was calculated on a token-by-token basis simply as the absolute value of the pre-conversation value of the acoustic correlate minus the
post-conversation value of the correlate. For example, the MS speaker in Pair 1 produced the /s/ in the word *casa* /ka.sa/ ‘house’ with a centre of gravity of 6175Hz in the pre-conversation production task, but with a centre of gravity of 6103Hz in the post-conversation production task. This represents an absolute change of |6175 – 6103| = 71.9Hz; the raw magnitude of the change for this speaker on the /s/ in this token was 71.9Hz. To determine the direction of this change, the pre-conversation measurement was compared against the mean COG for /s/ when it preceded /a/ for *all* of the speakers of the opposing dialect in the pre-conversation production task. That is, the direction of the change (convergence or divergence) was calculated relative to the entire other dialect, rather than relative to just the conversational partner. In this case, the mean COG for /s/ when /a/ follows produced by the BAS speakers in the pre-conversation production task was 6345Hz. This value is higher than the COG of /s/ in the MS speaker in Pair 1’s pre-conversation production of *casa*. When she produces this word in the post-conversation production task, her COG has fallen from 6175Hz to 6103Hz, moving away from the mean of the BAS speakers for /s/ followed by /a/ in the pre-conversation stage. Since the MS speaker has moved away the mean of the opposing dialect, her change of 71.9Hz is a divergence. Had she moved towards the mean of the opposing dialect, the change would be considered a convergence. This two-step method of calculating both the *magnitude* of the change and the *direction* of the change (relative to the opposing dialect) was repeated for each individual token produced by each individual speaker.

The direction of the change was defined in relation to the mean of the opposing dialect (taking into account the following vowel in the case of the fricatives), rather than just the mean of the conversational partner in order to avoid a situation where the relationship between the pre-conversation means of two speakers are reversed due to a particular conversational partner being an outlier on a given measure. For example, BAS speakers tend to have a higher COG for /s/ than MS speakers. However, it is possible that a particular BAS speaker could regularly produce /s/ with a low COG, causing his MS partner to have higher COG, thereby switching the expected relationship between the dialects within that pair. If the MS partner were then to lower his COG from pre- to post-conversation, this would appear to be convergence towards his BAS partner, but would certainly be considered divergence away from the BAS group and norm as a whole. In preliminary analysis, the direction of the effect was defined in both ways (relative to the partner, and relative to the entire opposing dialect) in order to determine how distinct the findings are for each definition. The difference between the findings was minimal, suggesting
that very few of the participants actually produced the sounds under investigation as real outliers relative to the other speakers of their dialect. Nevertheless, to avoid any impact of these outliers, the data presented in this thesis pertaining to the direction of the change from pre- to post-conversation are relative to the mean of the entire opposing dialect, rather than the mean of only the conversational partner.

5.3.2.1 A Note on Within-Category and Across-Category Accommodation

In §1, I provided an example of phonetic accommodation found in an episode of The Daily Show where the host, Jon Stewart, changes his front vowel /æ/ for a back vowel (something more like /ɑ/) in the final syllable of the words Afghanistan, Pakistan, and Taliban in response to his guest who produces the final syllables as the back vowel. This is an example of an accommodation in which one phoneme is substituted for another; the change is made across a category, resulting in a highly perceptible shift. Previous work on phonetic accommodation has found that convergences are not always (or even usually) so substantial. Instead, they are frequently gradient and, while statistically significant in many cases, often do not cause a shift from one phonemic category to another. Indeed, in the sentence immediately after producing /ɑ/ in place of /æ/, Jon Stewart changed back to his normal /æ/. A large shift that changes production of one phoneme into another introduces various potential problems in comprehension such as a collapsing of lexical contrasts. In addition, such a shift might fall within Trudgill (1986)’s notion of extra-strong salience (discussed in §3.1), where the D2 realization is strongly associated with the D2, discouraging speakers to take on that pronunciation.

In the current study I expect that the majority (if not all) of the changes that the participants make to their pre-conversation production will be gradient, within-category changes. This expectation is mostly related to the potential for changes within Difference #1 (BAS /s/ versus MS /θ/) and Difference #2 (BAS /ʃ/ versus MS /ʝ/) where a complete convergence would entail producing a segment that crosses a phonemic boundary. Of course, in the case of Difference #1, such a phonemic boundary only exists for the MS speakers, and in the case of Difference #2, only for the BAS speakers. Nevertheless, it is likely true that speakers of both dialects will be able to perceive the difference between the BAS and MS variants of both of
these dialectal differences\textsuperscript{10}. As such, shifting to take on the pronunciation of the opposing dialect for either of these two differences is somewhat unlikely. Instead, I expect gradient changes along the acoustic measures used to characterize the dialectal differences, for both the dialectal differences that reflect differences in phonemes as well as those that reflect within-category differences in articulation.

5.3.3 Standardizing the Magnitude of the Change

As we have seen, the magnitude of the change for the MS speaker in Pair 1 on \textit{casa} discussed above was 71.9Hz. The same speaker produced the segment /θ/ in the word \textit{azul} /a.'θul/ ‘blue’ (reflecting Difference #1: BAS /s/ v. MS /θ/) in the pre-conversation production task with a relative intensity of -21.7dB (the acoustic correlate used to characterize Difference #1). When she produced this word again post conversation, the relative intensity of the fricative was -23.9dB, giving an absolute raw change of 2.2dB. The change calculated for \textit{casa} (reflecting Difference #3) is in Hertz, while the change for \textit{azul} (reflecting Difference #1) is in decibels. How can we tell which of these changes is greater than the other? In order to allow comparison across the six dialectal differences, which use various acoustic correlates, the data needed to be standardized to one particular scale.

The method of standardization needed to maintain the relationship between the measurements taken on each token and the mean of that measurement for the opposite dialect. That is, the MS speaker in Pair 1 produced the /s/ in \textit{casa} in the pre-conversation word list with a lower COG than the speakers of BAS did, then produced it with a lower COG still in the post-conversation word list. In order to preserve this change as a divergence away from the BAS speakers after being standardized, the standardization process had to maintain the relationship between the pre-conversation, post-conversation, and opposing dialect values. In other words, if in the raw measurements a speaker diverges on a particular token, it must still be the case that the speaker diverges on that token after the magnitude of the change has been standardized. In addition, the method of standardization also had to take into the account the inherent variability of the six dialectal differences as produced by each speaker. This is important because a change made within a very stable distribution is likely a greater change than one made within a more variable distribution. To deal with this, I standardized the raw changes by dividing them by the

\textsuperscript{10} The results of the perception task confirming this are given in §6.2.
standard deviation of the pre-conversation production by each speaker and by each dialectal difference. For example, Table 11 shows some of the measurements taken for the MS speaker in Pair 1 for Difference #1, with the tokens alphabetically from azul ‘blue’ to zumo ‘juice’. The raw change is simply the absolute value of the difference between the pre- and post-conversation production of each token for this speaker.

Table 11: Standardizing the MS speaker in Pair 1’s changes made on Difference #1

<table>
<thead>
<tr>
<th>Token</th>
<th>Participant</th>
<th>Pre</th>
<th>Post</th>
<th>Raw Change</th>
<th>Standardized Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>azul</td>
<td>MS-1</td>
<td>-21.7</td>
<td>-23.9</td>
<td>2.2</td>
<td>2.2 / 2.52 = 0.8730</td>
</tr>
<tr>
<td>centro</td>
<td>MS-1</td>
<td>-24.6</td>
<td>-22</td>
<td>2.6</td>
<td>2.6 / 2.52 = 1.0317</td>
</tr>
<tr>
<td>zumo</td>
<td>MS-1</td>
<td>-24.3</td>
<td>-20.6</td>
<td>3.7</td>
<td>3.7 / 2.52 = 1.4683</td>
</tr>
</tbody>
</table>

To standardize the changes, I first calculated the standard deviation of this speaker’s pre-conversation production of /θ/. That is, I calculated the standard deviation of the relative intensity measurements taken on each of her pre-conversation productions of tokens containing /θ/ to get a value of how variable her production of that sound is. The standard deviation of the RI of her pre-conversation production of /θ/ was 2.52dB. I then divided the raw change made on each token by this standard deviation, giving the standardized changes in the rightmost column in Table 11. These standardized changes can now be directly compared with standardized changes made on other dialectal differences using different units, such as Difference #3, shown in Table 12. Here, the standard deviation of the MS speaker in Pair 1’s pre-conversation production of /s/ (measuring the centre of gravity of /s/) was 698Hz. I divided the raw changes made on each token containing Difference #3 by 698 to obtain the standardized change. Now we can compare the change that this speaker made on the /s/ in casa (0.1030) to the change she made on the /θ/ in azul (0.8730) to see that the change on azul is greater.

Table 12: Standardizing the MS speaker in Pair 1’s changes made on Difference #3

<table>
<thead>
<tr>
<th>Token</th>
<th>Participant</th>
<th>Pre</th>
<th>Post</th>
<th>Raw Change</th>
<th>Standardized Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>casa</td>
<td>MS-1</td>
<td>6175.1</td>
<td>6103.2</td>
<td>71.9</td>
<td>71.9 / 698 = 0.1030</td>
</tr>
<tr>
<td>casi</td>
<td>MS-1</td>
<td>5983.0</td>
<td>7263.3</td>
<td>1280.3</td>
<td>1280.3 / 698 = 1.8343</td>
</tr>
<tr>
<td>suma</td>
<td>MS-1</td>
<td>5176.4</td>
<td>6354.2</td>
<td>1177.8</td>
<td>1177.8 / 698 = 1.6874</td>
</tr>
</tbody>
</table>
This procedure was repeated for each speaker and each contrast. Standardizing by the standard deviation of each speaker’s pre-conversation production takes into account how variably each speaker produces the dialectal differences, such that small changes within very stable distributions will not be lost, while larger changes within more variable distributions will not be inflated. The data used in the statistical analysis of the magnitude of the change from pre- to post-conversation are always standardized, and data presented in graphics will also be standardized unless otherwise specified. In §6.1, the raw values of the changes are used because only the pre-conversation production is analyzed and only within each dialectal difference.

5.3.4 Mixed-Effects Modeling in R

The statistical data analysis (and the majority of the graphical analysis) in this thesis was conducted using R (R Development Core Team 2012). In the following chapter I present mixed-effects regression models built using the \textit{lme4} package (Bates, Maechler & Bolker 2011) and other graphical and analytical tools from the \textit{languageR} package (Baayen 2011), which are described in Baayen (2008), the \textit{sciplot} package (Morales & Murdoch 2011), and the \textit{multcomp} package (Hothorn, Bretz & Westfall 2008).

Mixed effects models incorporate both fixed effects, the variables that are chosen specifically for investigation in the study, and random effects, those that are sampled from larger populations. In the current study, fixed effects are variables such as DIALECT and GENDER, while the random effects are participant and token. There are several advantages to using a linear mixed-effects model over a repeated measures ANOVA. The first is that in an ANOVA, a single missing data point from one participant requires that the entire participant be dropped from the analysis. If missing data are common, this will substantially reduce the size of the available data for analysis. Mixed effects models, on the other hand, are robust to missing data (Quené & van den Bergh 2004), with only the missing data point being lost and any other data point associated with that participant being retained. The second advantage is that several of the assumptions necessary to use a repeated measures ANOVA are not required by mixed-effects models. Specifically, the assumptions of homogeneity of variance and sphericity are not made by mixed-effects models (Quené & van den Bergh 2004).

Relating to variation between participants, mixed-effects models do not require us to ignore interspeaker variation, rather the differences between participants can be modelled via
random effects. For example, in a study of response times to a stimulus it may be the case that some participants are faster responders than others. With ‘participant’ as a random effect, a mixed-effects model generates a coefficient for each participant that adjusts for this interspeaker variation. One advantage of this adjustment is that we do not need to average over participants to understand the behaviour of the group, instead the random intercepts control for the variation without having to lose power in the model. A second advantage is that the values of the random intercepts may be revealing of patterns between the participants reflecting some other factor. Furthermore, we can include by-speaker random slopes for an independent variable, which model individual sensitivity to that factor in terms of the dependant variable. Not only can random effects control for individual variation, they can also be used to explore patterns within the variability (Drager & Hay 2012).

To investigate the pattern of accommodation in terms of the magnitude of the change from pre- to post-conversation, a linear mixed effects model was built, while to investigate the pattern in terms of the direction of the change (a binary categorical variable: convergence or divergence per token) and the perceptual salience of the dialectal differences (a binary categorical variable: correct or incorrect per trial), logistic mixed effects models were built. Both types of models use the lmer() function in the lme4 package, with the logistic regression using a binomial distribution. To avoid over-fitting the models in this study, final models were compared to simpler models using likelihood ratio tests. The lme4 package does not provide p-values for linear mixed effects models due to difficulties in determining the correct degrees of freedom (Baayen 2008). In this study, p-values were calculated using the pvals.fnc() function in the languageR package (Baayen 2011) where possible, which generates Markov chain Monte Carlo (MCMC) p-values. MCMC sampling is not possible for models incorporating random correlation parameters, so for models where these parameters were warranted (as determined via likelihood ratio tests), p-values were estimated using the formula \(2 \times (1 - \text{pt(abs(t_statistic), #obs - #fixed_effects)})\), provided by Baayen (2008). The lme4 package does provide p-values for the z-statistics given in logistic regression, so the p-values provided by this package are used in the discussion of the direction of the change and of the perceptual salience of the dialectal differences, both of which have binary categorical dependant variables.
Chapter 6
Results

6.1 Production of Dialectal Differences Pre-Conversation

Before considering the pattern of phonetic convergence and the role of perceptual salience we must first examine the production patterns of the participants in this study from the baseline (pre-conversation) production task to determine how the six dialectal differences are produced before exposure to the contrasting dialect. As discussed in §4, the literature makes specific predictions about how these dialectal differences are expected to be produced and which acoustic measurements are likely to reflect the difference between the dialects. In this section I explore how the participants in the present study produce the dialectal differences in their baseline production task and how well the dialectal differences are expressed in terms of the acoustic correlates found in the literature. To verify that the realization of the dialectal differences was significantly different between the dialects, linear mixed-effects regression models were built and are described in the following subsections.

6.1.1 Dialectal Difference #1: BAS /s/ versus MS /θ/

Difference #1 reflects the realization of orthographic <c> (before <i> or <e>) and <z>. These graphemes are produced as the voiceless interdental fricative /θ/ in MS, and as the voiceless denti-alveolar fricative /s/ in BAS (Hualde 2005: 156). As discussed in §4, according to the literature, we expect that MS /θ/ will have a lower relative intensity (defined as the difference between the minimum intensity of the fricative and the maximum intensity of the following vowel) than BAS /s/ while a difference in centre of gravity might not be as reliable a cue (Hedrick & Ohde 1993).
In Figure 4, showing the distribution of the relative intensity (RI) of BAS /s/ and MS /θ/, we see that the mean RI for BAS /s/ (-17 dB) is higher than that of MS /θ/ (-21.8 dB), as expected. A linear mixed-effects model with RI in the pre-conversation session as the dependant variable, DIALECT as a fixed effect and token and subject as random effects finds that there is a significant difference between the mean RI of orthographic <c> (before <i> or <e>) and <z> of the BAS and MS speakers (t(677) = -3.10, p = 0.002). Figure 5 shows the mean RI values of BAS /s/ for the BAS participants and the mean RI values for MS /θ/ for the MS participants for each individual participant.

11 The error bars in this plot and all other plots represent the standard error of the mean.
This figure illustrates that while there is some variability between the participants, the general trend is for the MS participants to have a lower RI for their realization of orthographic <c> (before <i> or <e>) and <z> than the BAS participants. Additionally, we can use linear discriminant analysis to investigate how well the productions of these sounds are categorized as being produced by a speaker of BAS or MS using RI as the attribute to categorize them (Burns & Burns 2008: 591). A linear discriminant analysis with dialect as the grouping variable and relative intensity as the independent variable finds that overall 68.5% of the tokens containing Difference #1 are classified correctly on the basis of RI: 65.7% of BAS tokens and 71.3% of MS tokens were classified correctly. The addition of COG to the analysis along with RI made only a 1% increase in the proportion of tokens that were correctly classified, so, only RI will be used to differentiate between BAS /s/ and MS /θ/.

6.1.2 Dialectal Difference #2: BAS /ʃ/ versus MS /ʝ/

Difference #2 concerns the realization of orthographic <ll> and <y>, which are typically produced as the palatal approximant /ʝ/ in MS and as either the voiceless or voiced alveopalatal fricative [ʃ] or [ʒ] in BAS. In fact, the voiced realization [ʒ] was very uncommon, with only one speaker (the BAS speaker in Pair 7) producing it regularly. This speaker was the oldest of the BAS participants (born in 1958, 52 years old at the time of the experiment). Various researchers have noted that the use of the voiceless variant is more common in younger BAS speakers (Hualde 2005: 31; Chang 2008). Chang (2008) found that BAS speakers born after 1975 used
the voiceless variant almost exclusively. In the current study, only the BAS speaker in Pair 7 was born before 1975, so it is no surprise that he uses the voiced variant, while the others do not. As discussed in §4, relative intensity (RI) is used to reflect the difference in consonantal strength between the BAS fricative and MS approximant. According to previous studies incorporating consonantal strength, the weaker, more vowel-like MS /ʃ/ is expected to have a higher intensity relative to the following vowel than the BAS /ʃ/ (Balakrishnan et al. 1996; MacLeod 2010). Any voiced fricatives produced by the BAS speaker in pair 7 might be somewhat higher in RI than the voiceless counterparts as a result of being voiced, and therefore phonologically weaker. However, research suggests that differences in duration and glottal activity are the relevant acoustic correlates of voicing contrasts in fricatives, rather than RI (Stevens, Blumstein, Glicksman, Burton & Kurowski 1992), and neither of these acoustic measures are used to characterize Difference #2, so the impact of having voiced variants along with voiceless variants in the current analysis is likely minimal.

Figure 6: Difference #2 - Distribution of relative intensity in dB of BAS /ʃ/ and MS /ʃ/

Figure 6 shows the distribution of relative intensity values for these two sounds by dialect. We see that the mean RI for the BAS fricative is lower than that of the MS approximant, as expected. Furthermore, a linear-mixed effects model with pre-conversation RI as the dependant variable, DIALECT as a fixed effect, and token and subject as random effects confirms that this difference is significant ($t(738) = 5.328, p = 0.001$).
If we examine the mean RI for each individual participant in Figure 7, we see that there is certainly variability between the participants, but that there is a clear difference between the BAS group and MS group.

Furthermore, a linear discriminant analysis with dialect as the grouping variable and relative intensity as the independent variable finds that 76% of the tokens produced by all participants are classified correctly as being produced by a BAS speaker or an MS speaker. 76.7% of the BAS tokens and 75.4% of the MS tokens were classified correctly indicating that RI is a reliable acoustic correlate of the difference between BAS /ʃ/ and MS /ʝ/.

6.1.3 Dialectal Difference #3: BAS laminal [s] versus MS apical [s]

As noted in §4, the most common realization of /s/ in MS is the apico-alveolar [s], whereas the predominant realization in the rest of the Spanish-speaking world, including in Buenos Aires, is the lamino-alveolar [s] (Hualde 2005: 156; Piñeros 2009: 136), and this difference is characterized using centre of gravity, where we expect the BAS realization to have a higher COG than the MS realization. Figure 8 shows the distribution of centre of gravity values for /s/ by dialect. As expected, the BAS realization has a higher mean COG (5917Hz) than the corresponding MS realization (5448Hz) although this difference was found to be only
marginally significant via a linear-mixed effects model with pre-conversation COG as the dependant variable, DIALECT as a fixed effect, and token and subject as random effects ($t(417) = -1.256, p = 0.075$).

![Box plot showing the distribution of COG of /s/ in Hertz by dialect]

**Figure 8: Difference #3 - Distribution of centre of gravity of /s/ in Hertz by dialect**

Figure 9 shows the mean COG for /s/ for each individual participant and illustrates that the speakers vary considerably. Although the general pattern shows that COG is higher for BAS /s/ than MS /s/, this inter-speaker variability likely caused the difference not to be statistically significant.

![Bar graph showing mean COG of /s/ by individual participant]

**Figure 9: Mean centre of gravity in Hertz of /s/ by individual participant**
A linear discriminant analysis with dialect as the grouping variable and COG as the independent variable finds that overall 57.9% of the tokens containing Difference #3 are classified correctly on the basis of COG: 52.6% of BAS tokens and 63.2% of MS tokens were classified correctly. This accuracy is somewhat lower than that of the first two dialectal differences, which is not surprising given the lack of statistical significance of the difference in COG and the interspeaker variability. Including the other spectral moments (standard deviation, kurtosis, and skewness) and relative intensity did very little to improve the accuracy of the linear discriminant analysis. Furthermore, none of these variables are predicted to reflect the difference in articulation between BAS and MS /s/. As a result, only COG is used in the analysis.

Even though the difference across the two groups was only marginally statistically significant, this does not necessarily mean that COG is not the appropriate measure to distinguish between the BAS and MS realizations of /s/. Previous research has suggested that COG is a correlate of coronal fricative contrasts, as discussed in §4, with more anterior places of articulation having higher COG than more posterior realizations (Forrest et al. 1988; Jongman et al. 2000; Li et al. 2010). This prediction is borne out here, but the difference is not significant; however, the difference in articulation of /s/ that we are representing here by COG is not a phonemic contrast for speakers of either MS or BAS. Instead, the difference is only a dialect-specific variation in articulation of a fricative that has much in common between the two dialects. Phonemic contrasts between sibilants of such similar articulation are not common cross-linguistically (Hall 1997) but Toda, a Dravidian language spoken in India and one of the languages studied in Gordon et al. (2002), has a 4-way voiceless sibilant contrast between the dental /ʂ/, the alveolar /s/, the retroflex /ʂ/, and the postalveolar /ʃ/. In that study, the authors found that the fricative within this group with the most front articulation (the dental) had the highest COG, which was significantly higher than each of the other fricatives, including the alveolar /s/, suggesting that COG may help to distinguish these contrasts. The mean COG for the Toda dental /ʂ/ was 5027Hz, while the mean COG for the alveolar /s/ was 4529Hz, a difference of only 498Hz. Similarly, the difference in mean COG between BAS /s/ and MS /s/ is 468Hz. In the case of Toda, a difference of about 500Hz is enough to signal a phonemic contrast, at least partially. The difference between the mean COG of the dental and alveolar Toda fricatives was statistically significant, but there is likely still some overlap between the distributions. This finding suggests that the 468Hz difference between BAS and MS /s/ may be enough to also distinguish between the two. No details on the variability of the Toda fricatives
was given, but it may be that the distributions are less variable than the distributions of BAS and MS /s/, since the former two are involved in a phonological opposition, while the latter two are not.

In sum, although the difference in mean COG of BAS /s/ and MS /s/ was only marginally significant, the results of studies of languages that have phonemic contrasts between similar coronal sibilants suggest that COG is an appropriate cue for the discrimination of BAS and MS /s/.

6.1.4 Dialectal Difference #4: BAS Velar [x] versus MS Uvular [χ]

The velar fricative /x/ is produced with a more posterior place of articulation, often described as uvular, in MS, while the same fricative is produced at the velum in BAS. Since a more posterior place of articulation creates a longer oral cavity in front of the constriction, MS /x/ is expected to have a lower value of F2 at the onset of the vowel following /x/ (Onset F2) than BAS (Gordon et al. 2002; Ladefoged 2005: 42). As we can see in Figure 10, Onset F2 of BAS /x/ is slightly higher than that of the MS /x/, but the difference is small (BAS: mean 2767 mels, MS: 2686 mels) and a linear-mixed effects model with pre-conversation ONSET F2 as the dependant variable, DIALECT as a fixed effect, and token and subject as random effects found that the main effect of dialect was not significant (t(434) = -1.312, p = 0.16).

![Figure 10: Difference #4 - Distribution of F2 in mels at onset of vowel following /x/ by dialect](image-url)
If we consider the pattern of Onset F2 at the individual participant level, as in Figure 11, we can see that there is some variability between the participants, but that most are fairly consistent at around 2600 mels.

![Figure 11: Mean F2 in mels at onset of vowel following /x/ by individual participant](image)

As discussed in §4.4, there are very few studies on the acoustic correlates of velar/uvular place contrasts. The logic behind the use of Onset F2 to distinguish between the BAS and MS variants is that this particular acoustic correlate has found support in the literature previously (Gordon et al. 2002), while other potential correlates have not. In addition to the MS variant of /x/ being articulated farther back, it is also described as being more strident or turbulent than the BAS variant (Hualde 2005: 154), but it is not clear which acoustic measure could be used to quantify this quality. One possibility is relative intensity, which we might expect to be lower in the more turbulent MS variant; however, the acoustic analysis found that the mean RI was -19.1dB for MS /x/ and -18.7dB for BAS /x/, which are almost identical and not statistically significantly different. As noted by Mazzaro (2011), while there is no work on the centre of gravity of /x/, we might expect that fricatives articulated farther back in the oral cavity (in this case, MS /x/) would have a lower COG than those articulated farther front (BAS /x/), reflecting the pattern seen in coronal fricatives (Jongman et al. 2000). This was the case in the current study with the MS /x/ having a mean COG of 1810Hz, which is higher than the mean COG for BAS /x/ of 1674Hz. However, this difference was not statistically significant.
A linear discriminant analysis with dialect as the grouping variable and both Onset F2 and COG as the independent variables found that 58% of the tokens produced by all participants were classified correctly as being produced by a BAS speaker or an MS speaker, but this accuracy did not decrease when COG was removed from the analysis, suggesting that the effects of Onset F2 and COG in distinguishing between the two dialects are not consistent. As such, only Onset F2, which has more support in the literature for being a correlate of back fricative contrasts (Gordon et al. 2002) than COG, will be used in the analysis.

Similar to the lack of significant difference between the BAS and MS variants of /s/, which are two versions of the same phoneme, it is perhaps not a surprise that the Onset F2 values for the BAS and MS variants of /x/ do not differ significantly either.

6.1.5 Dialectal Difference #5: Diphthongization of Mid Vowel Sequences in BAS, but not in MS

In general, sequences of two vowels in which the first vowel is /e/ (/eo/ and /ea/ were included in the present study) are predicted to be syllabified in hiatus with each vowel occupying the nucleus of its own syllable (Quilis 1993). Difference #5 reflects the possibility that such sequences may be syllabified as a diphthong with the first vowel /e/ being realized as the glide [j] in BAS, but not in MS\(^\text{12}\). Previous studies examining the acoustic correlates of the diphthong/hiatus contrast have agreed that the most robust acoustic cue used to discriminate diphthongs and hiatus is duration: hiatus are typically longer than diphthongs (Aguilar 1999; Hualde & Prieto 2002; Face & Alvord 2004; MacLeod 2007, 2012). In addition, as investigated by Garrido (2007), if the vowel /e/ is realized as [j], we might also expect the raised variant to have a lower F1 value and higher F2 than the more faithful [e] (Lieberman 1977; Stevens 1998; Johnson 2003; Ladefoged 2005: 42).

To determine how well these acoustic correlates characterize BAS and MS productions of mid vowel sequences, we must examine the realization of both /iV/ sequences and /eV/ sequences in each dialect. In Figure 12 below we see that for the /iV/ sequences, which are expected always to be realized as a diphthong [jV], both dialects produce these sequences with a

\(^{12}\) To say that diphthongization of mid vowel sequences in MS is not possible is too strong a stance. As noted in §4.6, diphthongization of mid vowel sequences is possible in MS, but is predicted to be much less frequent since the process is stigmatized in that dialect, while it is not stigmatized in BAS (Hualde, Simonet & Torreira 2008).
very similar duration\textsuperscript{13} (BAS: 120ms, MS: 122ms). In contrast, the mid vowel /eV/ sequences show a difference between the two dialects, with the MS speakers producing the sequences with a mean duration of 141ms and the BAS producing them with a mean duration of 129ms.

A linear mixed effects model with VOCALIC SEQUENCE DURATION as the dependant variable, DIALECT and FIRST VOWEL (/i/ or /e/) as fixed effects, and token and subject as random effects found a significant main effect of DIALECT ($t(851) = 4.03, p < 0.001$) and a significant interaction between DIALECT and FIRST VOWEL ($t(851) = -3.138, p = 0.002$). The significance of the interaction term (with BAS as the reference level for DIALECT, and /e/ as the reference level for FIRST VOWEL) indicates that the MS speakers produce the /eV/ sequences significantly longer than do the BAS speakers, but that the durations of the /iV/ sequences do not differ by the dialects. Since hiatus tend to be longer than diphthongs, this suggests that the BAS speakers are producing these /eV/ sequences as diphthongs more often than the MS speakers.

Figure 13 shows the mean durations of the /eV/ sequences by individual participant. We can see that there is certainly inter-speaker variation, but that the overall tendency is for the MS speakers to produce longer /eV/ sequences than the BAS participants.

\textsuperscript{13} The durations of vocalic sequences in this and the following sections have been normalized according to the procedure in §5.3.1.4 in order to avoid the spurious effect of speech rate.
A linear discriminant analysis with dialect as the grouping variable and normalized duration of the /eV/ sequence as the independent variable found that overall 58.7% of the tokens were correctly classified as being produced by a BAS speaker or an MS speaker, with 60.6% of the BAS tokens and 56.8% of the MS tokens correctly classified.

Another expected correlate of Difference #5 is the value of F1 in the first vocoid. If the BAS speakers produce /e/ in an /eV/ sequence as [j], we expect the value of F1 in the first vowel to be lower for those speakers than for the MS participants. Figure 14 shows a very similar pattern to that found for mean normalized duration of the vocalic sequences. Again, the two dialects produce /iV/ sequences very similarly, this time with almost identical mean values of F1 (BAS: 1185 mels, MS: 1172 mels), but show a distinction in the /eV/ sequences. The MS participants produced the first vowel in the /eV/ sequences with a significantly higher F1 than the BAS participants, suggesting that the first vowel was produced in a lower position in the mouth by the MS participants than their BAS counterparts.
Figure 14: Difference #5 - Mean F1 in Mels of first vowel in a vowel sequence by sequence type and by dialect

A linear mixed effects model with VOCOIDF1 as the dependant variable, DIALECT and FIRST VOWEL (/i/ or /e/) as fixed effects, and token and subject as random effects found significant main effects of both DIALECT ($t(851) = 2.05, p = 0.041$) and FIRST VOWEL ($t(851) = -3.80, p = 0.0002$), as well as a significant interaction between DIALECT and FIRST VOWEL ($t(851) = -7.69, p < 0.001$). The significant main effect of FIRST VOWEL shows that both dialects realize the first vocoid in the /eV/ sequences with a higher F1 than the first vocoid in the /iV/ sequences, but the significance of the interaction term (with BAS as the reference level for DIALECT, and /e/ as the reference level for FIRST VOWEL) indicates that the MS speakers make a larger distinction in F1 between the two first vocoids in these sequences, meaning that the MS speakers produce a lower vocoid in the /eV/ sequences than the BAS speakers do.

If we consider the inter-speaker variability in F1 of the first vowel in the /eV/ sequences for each individual participant (Figure 15), we again find that while the speakers do vary, there is a general tendency for the MS participants to have higher mean F1 for /e/ than the BAS participants. A linear discriminant analysis with dialect as the grouping variable and F1 of the first vocoid as the independent variable finds that 61.9% of the tokens are correctly classified as being produced by either a BAS or an MS speaker, with 63.3% of BAS tokens and 60.5% of MS tokens correctly classified.
In addition to expecting the first vocoid in a vocalic sequence to have a lower F1 if it has been realized as the glide [j], we might also expect the vocoid to be realized with a higher F2 if it is a glide, as a result of being articulated somewhat farther front than [e] (Lieberman 1977; Stevens 1998; Johnson 2003; Ladefoged 2005: 42). If the BAS speakers realize underlying /eV/ sequences as a diphthong [jV], we would expect the F2 value of the first vocoid for their /eV/ and /iV/ sequences to be similar. Also, if the MS speakers do not diphthongize /eV/ sequences then we would expect the F2 of the first vocoid in their /iV/ sequences to have lower F2 than the first vocoid in /eV/ sequences. Figure 16 shows the mean F2 of the first vocoid in both the underlying /iV/ and /eV/ sequences by dialect.
We can see that for both dialects the value of F2 of the first vocoid in the sequence is higher in the /iV/ sequences than in the /eV/ sequences, as would be expected. In addition, the BAS speakers do show higher F2 values for the first vocoid in the /eV/ sequences compared to the MS speakers, which we would expect to see if the BAS speakers were realizing this vocoid as a high glide. However, a linear mixed effects model with VOCOIDF2 as the dependant variable, DIALECT and FIRST VOWEL (/i/ or /e/) as fixed effects, and token and subject as random effects found that only the main effect of FIRST VOWEL was significant (t(851) = 3.30, p = 0.001), but not the main effect of DIALECT or the interaction between DIALECT and FIRST VOWEL. The significance of the main effect of FIRST VOWEL shows that F2 is higher in the first vocoid of the /iV/ sequences than in the /eV/ sequences, but the lack of significance for the DIALECT or interaction terms shows that the BAS and MS speakers make the same distinction in F2 between the two sequence types. As such, F2 will not be useful in distinguishing between BAS and MS productions of mid vowel sequences and will not be discussed further.

The results for duration of the sequence and the value of F1 of the first vocoid in the sequence both suggest that the BAS participants were more likely to raise the /e/ in the /eV/ sequences to the glide [j], thereby diphthongizing the sequence. While previous studies have confirmed that both duration and the value of F1 are acoustic correlates to the production of
diphthongs and hiatus, none has considered the relative contribution that these correlates might make to the perception of the difference between the two syllabifications. Face & Alvord (2004) found that duration was a robust cue in the perception of the contrast between diphthongs and exceptional hiatus, but that difference is one between sequences that both involve high vocoids, such as between [ja] (a diphthong) and [i.'a] (an exceptional hiatus). In that case a difference in the value of F1 of the first vocoid would not be expected. In the situation involving mid vowel sequences, it is not clear to what extent the durational difference and the F1 difference contribute to the perception of the diphthongization. While both are present in production, it is not necessarily the case that both are used by listeners to perceive the difference or if both are used, it is not known how the two correlates might be weighted by listeners. However, a linear discriminant analysis using both duration of the sequence and F1 of the first vocoid increases the proportion of tokens that are correctly classified as being produced by a BAS speaker or an MS speaker from the proportions correctly classified by either duration alone or F1 alone. The combination of the two acoustic correlates correctly classifies 63.2% of tokens overall, and 61.9% of BAS tokens and 64.5% of MS tokens. Since the literature predicts that both of these acoustic correlates of the difference between a mid vowel sequence realized as a diphthong or a hiatus are important, and since the linear discriminant analysis improves as a result of including both, it is logical to attempt to use both correlates here.

As noted above, we do not know the weighting of these two correlates that listeners would use in judging a particular sequence as being realized as a diphthong or a hiatus, but we can generate a weighting based on the production of these two sequences as being realized by either a BAS or MS speaker. A logistic regression analysis was conducted to predict the dialect of the speaker who produced each vocalic sequence with duration of the sequence and F1 of the first vocoid as predictors. A test of the full model against a constant only model was statistically significant, indicating that the predictors as a set reliably distinguished between BAS and MS speakers (chi square = 40.505, p < 0.000, df = 2). The Wald criterion showed that both duration (Wald = 6.567, p = 0.01) and F1 (Wald = 25.534, p < 0.000) made significant contributions to classifying a given vocalic sequence as being produced by a BAS or MS speaker. Furthermore, the logistic regression provides logistic coefficients (β values) that can be used to generate an equation that predicts the dialect of speakers, weighting the two predictor variables to maximize discriminability. The regression calculated that the β values were 7.139 for duration and 0.004 for F1, along with a constant of -6.124. Using these values, we can generate the equation in
Figure 17, which gives the likelihood of a particular vocalic sequence with given duration and F1 of its first vocoid being produced by an MS speaker (Burns & Burns 2008: 573).

\[
\text{Composite value} = -6.124 + 7.139 \times \text{duration} + 0.004 \times \text{F1}
\]

Figure 17: Equation to predict dialect of a given vocalic sequence based on duration of the sequence and F1 of the first vocoid

This equation allows us to weight the relative contributions of duration of the sequence and F1 of the first vocoid to generate a single value that reflects both acoustic correlates. As noted above, this value does not necessarily reflect the weighting of these correlates that a listener would use to judge a particular production of a vocalic sequence as being produced by a BAS speaker (diphthong) or an MS speaker (hiatus), but it does allow us to incorporate both correlates, thereby improving the discriminability of the data. The values generated by the equation in Figure 17 will be used in the analysis of the pattern of phonetic accommodation as a reflection of how BAS speakers produce these mid vowels sequences, rather than relying on either duration of the sequence or F1 of the first vocoid.

6.1.6 Dialectal Difference #6: Exceptional Hiatus in MS, but not in BAS

As mentioned in §4.7, the general rule of syllabification predicts that a sequence of two vowels containing a high vowel that is unstressed will always be realized as a diphthong (Quilis 1993). If the high vowel is not stressed, following Quilis’ rule of syllabification, the expected realization of the vocalic sequence is the diphthong [je] instead of the hiatus [i.e]. Many studies have investigated the so-called ‘exceptional hiatus’ found in some varieties of Spanish, including Peninsular Spanish (subsuming Madrid Spanish), in which a hiatus is produced even when the high vowel is not stressed. Dialectal Difference #6 is the possibility of exceptional hiatus in MS, but not in BAS. The primary acoustic correlate of this difference is duration, according to the many results investigating this contrast in the literature (Aguilar 1999; Hualde & Prieto 2002; Face & Alvord 2004; Macleod 2007, 2012). The vocalic sequence in the MS pronunciation, if produced as an exceptional hiatus, should be longer than that in BAS, which is realized as a diphthong.
Figure 18 shows the mean normalized duration of sequences that do not typically allow a potential exceptional diphthong (such as *siendo* ['sjen.do] ‘being’) and those that do (such as *riendo* ‘laughing’, expected as ['rjen.do] in BAS, but potentially [ri.'en.do] in MS) by dialect. We can see that the mean durations for the diphthongs are almost identical between the dialects (MS: 135ms, BAS: 138ms), while those for the potential exceptional hiatus are longer for MS speakers (174ms) than for BAS speakers (164ms). A linear mixed effects model with normalized\textsuperscript{14} VOCALIC SEQUENCE DURATION as the dependant variable, DIALECT and SEQUENCE TYPE (expected diphthong or potential exceptional hiatus) as fixed effects (with BAS as the reference level for DIALECT, and diphthong as the reference level for SEQUENCE TYPE), and token and subject as random effects found a significant main effect of SEQUENCE TYPE (t(524) = 2.767, p = 0.006) as well as a significant interaction between DIALECT and SEQUENCE TYPE (t(524) = 3.89, p < 0.001). This significance of the main effect of SEQUENCE TYPE shows that both dialects produce the potential exceptional hiatus sequences with a longer duration than the expected diphthongs, while the significance of the interaction term indicates that the MS speakers produce a greater durational distinction between the two sequence types than the BAS speakers.

An interesting finding here is that in both dialects, the sequences that could potentially incorporate an exceptional hiatus are significantly longer than those that are predicted to always be realized as diphthongs, as evidenced by the significant main effect of SEQUENCE TYPE. Exceptional hiatus is generally described in the literature as a Peninsular Spanish phenomenon, but we have found here that as a group the BAS speakers produce these sequences with a greater duration than diphthongs, although not with as great a difference as the MS speakers. It is possible that the potential for exceptional hiatus is not as limited to Peninsular Spanish as previously thought. Another possibility is that the BAS speakers living in Madrid have already taken on the realization of these sequences as exceptional hiatus due to having spent time in contact with MS speakers.

\textsuperscript{14} These duration measurements were normalized according to the procedure given in §5.3.1.4 to avoid any effect of speech rate.
Figure 18: Difference #6 – Mean normalized duration of vocalic sequences that are typically syllabified as a diphthong versus those that could be syllabified as an exceptional hiatus by dialect

If we examine the mean duration of the vocalic sequences that could be realized as exceptional hiatus by each individual participant in Figure 19, we see that there is again much interspeaker variability, but that many of the MS speakers show longer durations than the BAS speakers. A linear discriminant analysis with dialect as the grouping variable and normalized duration of the vocalic sequence as the independent variable finds that 56.1% of the tokens are correctly classified as having been produced by a BAS or MS speaker on the basis of duration, with 59.5% of the BAS tokens and 52.7% of the MS tokens correctly classified.
6.1.7 Summary of Pre-Conversation Production of Dialectal Differences

The preceding subsections of §6.1 examined how each of the six dialectal differences were produced in the pre-conversation production task by providing the results of the acoustic analysis, focusing on the particular acoustic dimensions that were expected to characterize the various differences. For Difference #1 (BAS /s/ vs. MS /θ/) and Difference #2 (BAS /ʃ/ vs. MS /ʝ/), the acoustic dimension of relative intensity was expected to distinguish between the dialectal variants and it was shown that in both cases the mean RI of the BAS variant differed statistically significantly from that of the MS variant, indicating that RI is an appropriate measure to characterize these dialectal differences. For Difference #3 (BAS laminal [s] vs. MS apical [ʂ]), the acoustic dimension that was expected to distinguish between the BAS and MS variants was centre of gravity. As we saw in §6.1.3, there was a marginally statistically significant difference between the BAS and MS variants that may not have reached significance at the 0.05 level due to interspeaker variability. Similarly, Difference #4 (BAS velar [x] vs. MS uvular [χ]) was expected to be well distinguished using the value of the second formant at the onset of the following /x/, but as was discussed in §6.1.4, the difference in mean F2 was not statistically significant between the two dialects. In the cases of Differences #3 and #4, the less robust difference in the acoustic measurement may be due to these differences being within-
category differences in articulation, rather than between-category differences that span a phoneme boundary, as in Differences #1 and #2. For Difference #5 (diphthongization of mid vowel sequences in BAS, but not in MS), §6.1.5 showed that a composite measure combining the weighted values of duration of the vocalic sequence and the value of F1 of the first vocoid in the sequence characterized the difference between the two dialects well. Similarly, for Difference #6 (potential exceptional hiatus in MS, but not in BAS), the difference was distinguished using the duration of the vocalic sequence, as the literature predicts (Aguilar 1999; Hualde & Prieto 2002; Face & Alvord 2004; MacLeod 2007, 2012).

It was expected that the MS speakers would show a greater difference in duration between the sequences where diphthongs are expected and the sequences where exceptional hiatus are possible than the BAS speakers would. This was the result that was found, but the BAS speakers also produced the sequences in words where exceptional hiatus are possible as longer than where diphthongs are expected, where we might have expected them to show no difference in duration between the two sequence types. As mentioned in §6.1.6, it is possible that the BAS speakers may already have come to converge towards the MS pronunciation of these exceptional hiatus as a result of having spent time living in Madrid. In fact, this is possible for any of the six dialectal differences being considered in this study. All of the participants had spent time in Madrid at the time of their participation in the experiment, ranging from 1 month to 8 years. To the extent that the BAS speakers have already converged towards the MS norms, the characterization of the dialectal differences via the acoustic measures given in this chapter will be underrepresented. One way to determine whether the BAS participants have made perceptible shifts in pronunciation would be to conduct an accent rating task in which native speakers of BAS judge snippets of speech taken from the BAS participants in this study, interspersed with samples of BAS speakers who have not left Buenos Aires, and provide accent ratings. This methodology has been used previously to assess phonetic accommodation, such as in Munro et al. (1999)’s study of the speech of Canadians living in Alabama. Determining how much the BAS speakers’ pronunciation in the pre-conversation production task differs from BAS speakers who have not left Buenos Aires, and thereby evaluating how much convergence towards MS norms had already taken place when the BAS participants took part in the current study, will be left to future research.
6.1.8 Repetition of Opposing Dialect Variants

As explained in §5.2.2, part of the procedure of the experiment involved the participants performing a rapid shadowing task (Goldinger 1998) in which they heard various stimuli, each of which contained exactly one of the dialectal differences, and then repeated the token immediately. The repetition task included stimuli containing both the BAS and the MS variants of each dialectal difference. The purpose was to ensure that the participants were actually able to reproduce the opposing dialect variants. If the participants were not able to produce these variants, then any lack of convergence found in the comparison of the pre- and post-conversation production tasks could potentially be the result of an inability to articulate the opposing dialect variant, as opposed to an effect of perceptual salience or some other factor.

To ensure that the participants were able to articulate the sounds of the opposing dialect, the recordings made during the repetition task were carefully evaluated perceptually. After hearing the production of each token, I made a judgement about which token had been produced and then checked the list of tokens for that participant (since the order was randomized each time) to verify that my perception of the token matched the target. There were only a handful of discrepancies and they showed no pattern across the dialectal differences or the dialects. As such, I conclude that the participants do have the necessary ability to produce the sounds of the opposing dialect.

6.2 Perceptual Salience of the Dialectal Differences

Perceptual salience has been defined in various ways in previous studies. As discussed in §3, many of these definitions are potentially problematic in that they do not assess the perception of salience directly and assume that all speakers of a given dialect will perceive a dialectal difference in the same way. That is, they assume that a dialectal difference will be equally salient for all speakers. In the present study perceptual salience of the six dialectal differences between BAS and MS was established via a perception task, which was discussed in §5.2.3, following the insights and methodology of Graff et al. (1986), Fridland et al. (2004), Thomas & Reaser (2004), and Torbert (2004, 2010). The current section details the results of the perception task to show the variation in perceptual salience of the six differences.

To briefly summarize the procedure of the perceptual task, the participants saw the orthographic form of a stimulus on the computer screen and heard two pronunciations of that
stimulus. Their task was to decide which of the two repetitions reflected the pronunciation norms of their native dialect area. There were 4 possible combinations of pronunciation of the two repetitions:

1. Both repetitions were produced by speakers of MS
2. Both repetitions were produced by speakers of BAS
3. The first repetition was produced by a speaker of MS and the second by a speaker of BAS
4. The first repetition was produced by a speaker of BAS and the second by a speaker of MS

The participants saw the orthographic form of the stimulus on the screen, heard the two repetitions and then had 5 possible responses:

1. that only the first repetition sounded like their native dialect
2. that only the second repetition sounded like their native dialect
3. that both repetitions sounded like their native dialect
4. that neither repetition sounded like their native dialect
5. that they did not know whether either repetition sounded like their native dialect

6.2.1 Control Trials

The stimuli in the perception task were sequences of sounds (as described in §5.2.2.1) that contained exactly one of the six dialectal differences along with control trials that included none of the differences. Examples of the control trials are [ba], [de], or [tu]. We begin with the results of the controls. Since these trials did not include any of the six dialectal differences (or any other known difference that could be a cue to dialect), it was expected that the participants would not be able to reliably identify which dialect the control stimuli came from. If the participants were able to identify the control trials at above chance level, this would suggest that there was some information in the vowels or the realization of consonants not generally described as containing dialectal variation that the participants were using to identify the control trials. In that case, we would not be able to conclude that any above-chance results in the test trials were the result of the presence of one of the six dialectal differences. On the other hand, if the participants respond to the control trials at chance levels, this would suggest that there is no information outside of the six dialectal differences that contributes to their ability to identify the dialects. Figure 20 below shows the percentage of responses by accuracy (‘incorrect’, ‘correct’, or ‘do not know’) by dialect. We see that for each dialect the percentage of correct responses is very close to 20%.
Figure 20: Percentage of responses to the control trials by participants’ dialect and accuracy

Since there were five possible responses in the perception task, only one of which was correct, chance is 20%. The BAS participants responded correctly to 21% and the MS participants responded correctly to 23% of the control trials. One-sample t-tests confirm that these response rates are not statistically significantly above 20% (MS: $t=0.849$, $df=10$, $p=0.416$, BAS: $t=0.363$, $df=10$, $p=0.724$). In addition, there is no significant difference between the accuracy rates of the BAS and MS speakers on the control trials, since a logistic mixed effects model found no significant main effect of dialect. However, a close inspection of the accuracy rates on the control trials found that the MS speakers responded accurately statistically significantly more often on the control trials [ma] and [na] (the only trials to include nasal consonants) than the BAS speakers did (33% of the time for MS speakers versus 11% for BAS speakers), whereas no such difference existed between the remaining (non-nasal) trials (22% accuracy for BAS speakers, 24% for MS speakers). This asymmetry in the accuracy of the control trials involving nasal consonants was particularly pronounced when one of the repetitions was produced by a BAS speaker and the other by an MS speaker. In this case, the MS speakers responded correctly to the control trials involving nasal consonants a whopping 63% of the time, compared to only 5% of the time for BAS speakers. It is not clear what acoustic characteristic of these nasal trials the MS speakers are using to generate such high accuracy rates, or whether this characteristic is

---

15 The one-sample t-test assumes that the data are normally distributed. This assumption is met for each dialect: MS: Shapiro-Wilk statistic 0.964, $df = 11$, $Sig = 0.815$; BAS: Shapiro-Wilk statistic 0.959, $df = 11$, $Sig = 0.764$. 
something that would allow them to distinguish between BAS and MS voices generally or if it is something specific to the recordings of these tokens in the current study. Nevertheless, the asymmetry in the accuracy of the control trials between the dialects is limited to [ma] and [na] and none of the test trials involving the 6 dialectal differences include nasals. So while the MS speakers may be making use of some acoustic characteristic of these sounds to identify the dialect of the model speaker, this is unlikely to affect the accuracy rates in the test trials.

Overall then, the control trials are responded to at chance levels. Focusing on the results of the control trials by individual participants, we see in Figure 21 that there is some variability in the percentage of correct responses between participants. The proportion of correct responses ranges from 2% to 39%, but most are either slightly above or slightly below chance level of 20%.

![Figure 21: Percentage of correct responses to the control trials by individuals](image)

This result shows that the overall the participants are not able to reliably identify the dialect of the control trials in the absence of one of the six dialectal differences under investigation. This indicates that any improvement in accuracy in the test trials will be the result of the presence of one of the dialectal differences. In this way, the percentage of test trials containing each dialectal difference that are responded to correctly is taken as a reflection of the perceptual salience of that particular dialectal difference.
6.2.2 Perceptual Salience of the Dialectal Differences

We turn now to the results of the test trials, which are those that include the six dialectal differences of interest here. Figure 22 illustrates the percentage of trials that were correctly responded to (referred to as %-Correct) for each of the six dialectal differences by dialect of the participants. The percentage of trials correctly responded to (accuracy rate) is taken to reflect the perceptual salience of each dialectal difference. In general, Figure 22 suggests that there are two groups of dialectal differences based on perceptual salience: the highly perceptually salient differences on the left-hand side of the chart, and the less salient differences on the right-hand side of the chart. Those on the left are Differences #1 and #2, which were predicted to be very salient.

![Figure 22: Percentage of correct responses by dialectal difference by dialect of the participants](image)

Table 13 below gives the exact accuracy rates for each of the dialectal differences by dialect of the participants.
Table 13: Mean %-Correct by dialectal difference and dialect, and overall mean

<table>
<thead>
<tr>
<th>Dialectal Difference</th>
<th>BAS</th>
<th>MS</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1: BAS /s/ vs. MS /θ/</td>
<td>75%</td>
<td>88%</td>
<td>82%</td>
</tr>
<tr>
<td>#2: BAS /ʃ/ vs. MS /ʝ/</td>
<td>82%</td>
<td>82%</td>
<td>82%</td>
</tr>
<tr>
<td>#3: BAS [s] vs. MS [s]</td>
<td>25%</td>
<td>22%</td>
<td>24%</td>
</tr>
<tr>
<td>#4: BAS [x] vs. MS [χ]</td>
<td>31%</td>
<td>39%</td>
<td>35%</td>
</tr>
<tr>
<td>#5: BAS diphthongization of midV sequences</td>
<td>28%</td>
<td>51%</td>
<td>40%</td>
</tr>
<tr>
<td>#6: MS exceptional hiatus</td>
<td>23%</td>
<td>34%</td>
<td>29%</td>
</tr>
</tbody>
</table>

To determine whether the six dialectal differences varied in accuracy rate (that is, whether they differed in perceptual salience) and whether the BAS and MS speakers differed in how perceptually salient they found the six differences, a logistic mixed effects regression model was built with a binomial distribution, which captures the binary, categorical nature of the dependant variable, ACCURACY (correct or incorrect), which was assessed for each trial that each participant responded to. The random effects were subject and stimulus. The fixed effects were DIALECTAL DIFFERENCE, DIALECT of the participant, and PAIR ORDER (the order in which the pair of stimuli were presented on each trial). Although there were 4 possible orderings in the experiment (MS-MS, BAS-BAS, BAS-MS, MS-BAS), preliminary analysis found that there was no difference in accuracy between the BAS-MS and MS-BAS orderings, so these two were collapsed into one level called “DIFFERENT”, since each of these two orderings involves presentation of a stimulus from each dialect, whereas MS-MS and BAS-BAS involve presentation of the same dialect twice.

All of the fixed effect predictor variables were categorical. The lmer() function selects one of the levels of a categorical variable as the reference level. The reference level for DIALECTAL DIFFERENCE is Control, the reference level for DIALECT OF PARTICIPANT is BAS, and the reference level for PAIR ORDER is DIFFERENT. The categorical variables were sum-coded and centred at 0 to ensure that the coefficient for each level was compared to the mean of the reference level, but that the overall intercept of the model was calculated as the grand mean of all of the levels.

The output of the model, with the control trials as the reference level for DIALECTAL DIFFERENCE, is presented in Table 14 below, giving the β estimates, standard error, z-values and

---

16 The results for the perception task, along with the raw changes for each dialectal difference, for each participant are provided in Appendix E.
p-values for each term. Main effects for DIALECTAL DIFFERENCE, DIALECT, and PAIR ORDER were included, along with an interaction between DIALECT of participant and DIALECTAL DIFFERENCE and an interaction between DIALECT of participant and PAIR ORDER.

Table 14: Model summary: logistic mixed effects model with Accuracy as dependant
Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coef. Est.</th>
<th>Std. Error</th>
<th>z-value</th>
<th>p-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.89110</td>
<td>0.19403</td>
<td>-4.593</td>
<td>&lt;0.001</td>
<td>***</td>
</tr>
<tr>
<td>DiaDiff1_θ-s</td>
<td>2.64632</td>
<td>0.31562</td>
<td>8.385</td>
<td>&lt;0.001</td>
<td>***</td>
</tr>
<tr>
<td>DiaDiff2_j-ʃ</td>
<td>3.00539</td>
<td>0.21943</td>
<td>13.696</td>
<td>&lt;0.001</td>
<td>***</td>
</tr>
<tr>
<td>DiaDiff3_s</td>
<td>0.31366</td>
<td>0.31127</td>
<td>1.008</td>
<td>0.314</td>
<td></td>
</tr>
<tr>
<td>DiaDiff4_x</td>
<td>0.60915</td>
<td>0.30024</td>
<td>2.029</td>
<td>0.042</td>
<td>*</td>
</tr>
<tr>
<td>DiaDiff5_midVs</td>
<td>0.34348</td>
<td>0.27457</td>
<td>1.247</td>
<td>0.212</td>
<td></td>
</tr>
<tr>
<td>DiaDiff6_Hiatus</td>
<td>0.14709</td>
<td>0.28280</td>
<td>0.520</td>
<td>0.603</td>
<td></td>
</tr>
<tr>
<td>Dialect_MS</td>
<td>0.05517</td>
<td>0.24489</td>
<td>0.225</td>
<td>0.822</td>
<td></td>
</tr>
<tr>
<td>DiaDiff1_θ-s_MS</td>
<td>0.98240</td>
<td>0.44738</td>
<td>2.196</td>
<td>0.028</td>
<td>*</td>
</tr>
<tr>
<td>DiaDiff2_j-ʃ_MS</td>
<td>0.09466</td>
<td>0.27575</td>
<td>0.343</td>
<td>0.732</td>
<td></td>
</tr>
<tr>
<td>DiaDiff3_s_MS</td>
<td>-0.41204</td>
<td>0.39424</td>
<td>-1.045</td>
<td>0.296</td>
<td></td>
</tr>
<tr>
<td>DiaDiff4_x_MS</td>
<td>0.14509</td>
<td>0.35724</td>
<td>0.406</td>
<td>0.685</td>
<td></td>
</tr>
<tr>
<td>DiaDiff5_midVs_MS</td>
<td>0.94237</td>
<td>0.32425</td>
<td>2.906</td>
<td>0.004</td>
<td>**</td>
</tr>
<tr>
<td>DiaDiff6_Hiatus_MS</td>
<td>0.36970</td>
<td>0.33231</td>
<td>1.113</td>
<td>0.266</td>
<td></td>
</tr>
<tr>
<td>PairOrder_BASBAS</td>
<td>-0.44603</td>
<td>0.18360</td>
<td>-2.429</td>
<td>0.015</td>
<td>*</td>
</tr>
<tr>
<td>PairOrder_MSMS</td>
<td>-1.25339</td>
<td>0.20104</td>
<td>-6.235</td>
<td>&lt;0.001</td>
<td>***</td>
</tr>
<tr>
<td>Dialect_MS*PairOrder_BASBAS</td>
<td>-0.83563</td>
<td>0.23158</td>
<td>-3.608</td>
<td>&lt;0.001</td>
<td>***</td>
</tr>
<tr>
<td>Dialect_MS*PairOrder_MSMS</td>
<td>1.21323</td>
<td>0.23794</td>
<td>5.099</td>
<td>&lt;0.001</td>
<td>***</td>
</tr>
</tbody>
</table>

Tukey post-hoc comparisons on the 7 levels of DIALECTAL DIFFERENCE (6 differences + control trials) were run using the glht() function in the multcomp package (Hothorn, Bretz & Westfall 2008), once for each of the two dialects. The purpose of the post-hoc comparisons is to determine if there are statistically significant differences in accuracy rates (salience) between the six dialectal differences within each dialect\(^{17}\). The results show that for the BAS speakers there are essentially two levels of salience within the 6 dialectal differences: high (encompassing Differences #1 and #2, which did not differ from each other in terms of salience) and low (encompassing all other dialectal differences, which did not differ from each other, but were significantly lower in salience than Differences #1 and #2). Table 15 below shows the post-hoc

\(^{17}\) The significant interaction between dialect and some of the dialectal differences allows post-hoc tests to be run.
comparisons for the BAS speakers. Differences #1 and #2 are significantly higher than all the others. The other differences are not significantly higher than the controls.

Table 15: Results of post-hoc tests for salience of the dialectal differences for BAS speakers

<table>
<thead>
<tr>
<th></th>
<th>Diff #1</th>
<th>Diff #2</th>
<th>Diff #3</th>
<th>Diff #4</th>
<th>Diff #5</th>
<th>Diff #6</th>
<th>control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diff #1 (BAS /s/ vs. MS /θ/)</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Diff #2 (BAS /ʃ/ vs. MS /ʝ/)</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Diff #3 (BAS [s] vs. MS [ɕ])</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Diff #4 (BAS [x] vs. MS [χ])</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Diff #5 (BAS midV diph.)</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Diff #6 (MS excep. hiatus)</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>control</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
</tbody>
</table>

These results show that for the BAS speakers there are two groups of dialectal differences defined by their salience, as predicted above, which are illustrated in Figure 23.

Low Salience  High Salience

Difference #3
Difference #4
Difference #5
Difference #6

Difference #1
Difference #2

Figure 23: Illustration of the perceptual salience of the 6 dialectal differences for the BAS speakers

The result for the MS speakers was similar, except that Difference #5 (midVs) fell in between Differences #1 and #2 (high salience) and the other differences, giving a mid level of salience for the MS speakers consisting of only Difference #5. Table 16 below shows the results of the post-hoc comparisons for the MS speakers. Again, Differences #1 and #2 are higher than all the others. Differences #3, #4 and #6 do not differ from the controls, but Difference #5 falls in between the high and low salience differences.
Table 16: Results of post-hoc tests for salience of the dialectal differences for MS speakers

<table>
<thead>
<tr>
<th></th>
<th>Diff #1</th>
<th>Diff #2</th>
<th>Diff #3</th>
<th>Diff #4</th>
<th>Diff #5</th>
<th>Diff #6</th>
<th>control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diff #1 (BAS /s/ vs. MS /θ/)</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Diff #2 (BAS /ʃ/ vs. MS /ʝ/)</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Diff #3 (BAS [s] vs. MS [g])</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff #4 (BAS [x] vs. MS [χ])</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diff #5 (BAS midV diph.)</td>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Diff #6 (MS excep. hiatus)</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
</tbody>
</table>

This result shows that there were three groups of salience for the MS speakers. These groups are illustrated in Figure 24 below.

Low Salience ————> Mid Salience ————> High Salience

Difference #3      Difference #5      Difference #1
Difference #4
Difference #6

Figure 24: Illustration of the perceptual salience of the 6 dialectal differences for the MS speakers

These findings show that there is variability in how perceptually salient the six dialectal differences are, which allows an investigation of the effect of perceptual salience on phonetic accommodation. If all the six differences were responded to correctly at similar rates, then we would not be able to investigate the role of perceptual salience in the pattern of accommodation.

The mean accuracy rate across all dialectal differences for BAS speakers was 40%, while for MS speakers it was 46%. Although the main effect of DIALECT is not statistically significant, the post-hoc tests found that Difference #5 was more salient than the low salience differences for the MS speakers but not for the BAS speakers. The difference in accuracy for Difference #5 is 23% (28% for BAS speakers and 51% for MS speakers). The MS speakers’ better performance may stem from the fact that diphthongization of mid vowel sequences is stigmatized in Peninsular Spanish but not in Argentine Spanish (Hualde et al. 2008).
The model in Table 14 also shows a significant main effect of **PAIR ORDER**, with the BAS-BAS and MS-MS orders compared to the reference level of **DIFFERENT** (encompassing both MS-BAS and BAS-MS orders). The coefficients for the BAS-BAS and MS-MS levels of **PAIR ORDER** are negative, indicating that accuracy rates on the trials involving these two orders were lower than those where a stimulus from each of the dialects was presented. This outcome shows that where there was a contrast to be perceived, the participants were more able to accurately identify the stimulus from their own dialect, versus when both stimuli were produced by speakers of the same dialect. In addition, there was a significant interaction between **DIALECT** of the participant and **PAIR ORDER**, where MS participants performed more accurately on the MS-MS trials than the BAS participants, and the BAS participants performed more accurately on the BAS-BAS trials than the MS participants. This interaction shows that the participants are more accurate in identifying stimuli that were produced by a speaker of their own dialect than identifying those that were produced by a speaker of the opposing dialect.

### 6.2.3 The Unexpected Salience of Difference #3

In §4, the six dialectal differences were presented and it was predicted that Differences #1 and #2 would be the most salient, since they both contain variants that are strongly associated with one of the dialects and reflect differences that cross phoneme boundaries (or contain a variant not found in one of the dialects). The results of the perception task confirm that these two differences are the most salient; their accuracy rates were significantly higher than all the other differences for both dialects. In §4, it was also predicted that the least salient differences would be those involving the treatment of vocalic sequences (Differences #5 and #6), while Differences #3 and #4 would have a medium level of salience. As we have seen, for the BAS speakers there was no distinction between any of these four dialectal differences in terms of salience. All of them were found to be significantly less salient than Differences #1 and #2, but no more so than the control trials. For the MS speakers, Difference #5, which was predicted to be of low salience, was actually found to be of medium salience.

The most surprising result is the very low accuracy rate with which the participants responded to the trials involving Difference #3: 25% for BAS speakers and 22% for MS speakers, neither of which was higher than chance. This finding is surprising since the apical realization of MS /s/ was one of the most commonly imitated features of Peninsular Spanish in Reiter (2004)’s study of Spanish accent imitation, which according to Trudgill’s criteria for
salience should render the difference salient (Trudgill 1986: 11). In addition, the difference in realization of the fricative /s/ was cited by six of the participants in the present study in the questionnaire as one of the noticeable differences between MS and BAS, suggesting that (at least some of) the participants are aware of the difference. If they are, then why were they unable to accurately identify Difference #3 in the perception task, causing it to appear to be one of the least salient of the six dialectal differences?

One possibility is that the perception stimuli that the participants used to make their judgments, which were produced by BAS and MS speakers, were not significantly different from each other. However, as mentioned in §5.2.2.1, the perception stimuli were analyzed acoustically and statistical tests confirmed that they differed from each other in the expected direction, with the mean COG for /s/ in the perception stimuli being 7718Hz for the BAS speakers and 5312Hz for the MS speakers. This difference was statistically significant (the results of these tests are given in Appendix D), so it seems that the lack of strong salience is not due to a lack of distinction in the perception stimuli.

One possible reason for the finding of low perceptual salience for /s/ is that this dialectal difference is actually not as perceptually salient as its involvement in imitation (Reiter 2004) and the fact that participants mentioned it in the questionnaire would suggest. It is possible that Spanish speakers know that this difference exists not because of their actual ability to perceive the difference but because they are aware of the stereotype that /s/ in Peninsular Spanish dialects is different from the /s/ in many other dialects. As noted by Hickey (2000), “stereotypical features are not usually determined afresh by each generation, still less by each speaker. Rather they are part of the inherited knowledge of features which are putatively typical for a certain variety” (Hickey 2000: 58). As such, knowledge of stereotyped linguistic variables “does not necessarily rest on direct experience of the variety which is being alluded to. This fact explains why stereotypes are so frequently off the mark” (Hickey 2000: 58). Torbert (2004), in his discussion of isolating the perceptual salience of individual linguistic variables, notes that simply asking speakers about their perceptions of the differences between two varieties of their language can be inaccurate, and points to the notion of ‘drawl’ as an example, saying “The term ‘Southern drawl’ comes across as a rehearsed or formulaic response when people give opinions on what makes Southern speech distinct from Northern varieties, but no one is really sure what drawl means in precise, linguistic terms” (Torbert 2004: 65). It may be in the same way that
Spanish speakers know there is some difference between how /s/ is articulated between BAS and MS, but only because it is something that people talk about, and when actually tested on their perceptions, they are unable to accurately identify /s/ produced by speakers of MS and BAS.

The above is a possibility; however, a more likely reason stems from an interaction between the salience of the difference in realization of /s/ across the dialects and the quality of the vowel following the /s/. As discussed in §5.2.2.1, the stimuli in the perception task to determine the perceptual salience of Difference #3 were sa [sa], asa ['a.sa], asá [a.'sa]. Since these stimuli only included one following vowel (/a/), the experiment was limited to measuring how perceptually salient /s/ was for the participants when /a/ followed. It is possible that the perceptual salience of the difference in realization of /s/ between the two dialects is lower in this particular context than in other contexts. If so, then the value of %-Correct in the perception task will underestimate the perceptual salience of Difference #3. As we saw in §6.1.3, in general MS /s/ has a lower COG than BAS /s/. In addition, there is a significant effect of following vowel quality on the COG of /s/. As we can see in Figure 25, when /s/ precedes /o/ or /u/ the centre of gravity of /s/ is lower than when it precedes another vowel. This effect is found for both dialects.

![Figure 25: Mean centre of gravity in Hertz for pre-conversation MS and BAS /s/ by following vowel](image)

Focusing on the COG for /s/ when it precedes /a/ (as it does for all of the stimuli in the perception task), we see that the difference in COG in this context between the two dialects is actually larger than when other vowels follow. This might seem to suggest that if the participants could perceive a dialectal difference in any vowel context it should be when /s/
However, we can also see from Figure 25 that the mean COG for MS /s/ preceding /a/ is almost exactly the same as the COG for BAS /s/ when it precedes /o/ or /u/. This means that, although the raw difference in COG between MS and BAS /s/ in front of /a/ is larger than in other vowel contexts, the COG for MS /s/ in this context still falls within the range of possible COG values for BAS /s/. In contrast, when MS /s/ precedes /o/ or /u/ the COG is lower than even the lowest possible values for BAS /s/. This may indicate that the difference between the two dialectal pronunciations of /s/ may be more evident, more salient, when the following vowel is /o/ or /u/ than when it is /a/. Supporting this is the fact that, while segmenting the data files, I made note of any productions that seemed to perfectly exemplify the dialectal difference they contained. For MS speakers producing /s/, the tokens that I noticed were *suma* ['su.ma] ‘sum’ (twice), *solo* ['so.lo] ‘alone’ (twice), *consumo* [con.'su.mo] ‘consumption’, *resultó* [re.sul.'to] ‘to work out.3sg.past’, *sujeto* [su.'je.to] ‘subject’, *sonido* [so.'ni.do] ‘sound’, and *siendo* ['sjen.do] ‘being’, all of which, except for *siendo*, have /o/ or /u/ following the /s/. No tokens with /a/ after the /s/ were noticed as being especially reflective of the MS realization of /s/18.

The above may help to explain why the difference in articulation of /s/ between the dialects was not found to be perceptually salient to the BAS speakers, but in the case of the MS speakers, the COG of BAS /s/ before /a/ is far higher than the COG of MS /s/ before /a/, which is where COG is highest in MS. In the same way that MS /s/ before /o/ and /u/ might be more salient to BAS speakers because the COG of /s/ in this context falls well below the means expected within the distribution of BAS /s/, we might have expected that since the COG of BAS /s/ before /a/ falls well above the means expected for MS /s/ that the MS speakers would be more accurate in perceiving the dialectal difference in this context. However, as we have seen, there was no statistically significant difference in accuracy rate between the two dialects. It is possible that differences in COG at the higher frequencies are not as salient as differences in the lower frequencies. That is, perhaps the difference between an alveolar and a front alveolar voiceless fricative is not as salient as the difference between an alveolar and a more posterior alveolar fricative. If this is the case, it may explain why the MS speakers were also fairly

18 Although I am not a native speaker of Spanish, the difference in articulation between BAS [s] and MS [ś] is certainly perceptible to me, likely because the difference approximates my native English /s/-/ʃ/ contrast, which relates to an idea that will be explored in §7.2.
unsuccessful in perceiving the difference between BAS /s/ and MS /s/. Of course, without testing the perceptions of speakers regarding the discriminability of /s/ at varying points on the COG continuum, it is difficult to say.

If the difference in realization of /s/ between BAS and MS is more salient when /o/ or /u/ follows than when /a/ follows, then the measure of perceptual salience for Difference #3 according to the perception task in the present study will be underestimated. If the salience of Difference #3 is underestimated, then any attempt to examine the pattern of accommodation in terms of the perceptual salience of the dialectal differences will be distorted. In light of the above pattern relating to the effect of the vowel following /s/, both in terms of COG and in terms of the tokens that were picked out during data analysis, it seems likely that the salience of Difference #3 has been underestimated here, but it is not clear by how much. Any attempt to correct the accuracy rate (without rerunning the perception task with new stimuli) would be purely speculative. As such, Difference #3 will not be included in the analysis of the role of perceptual salience on the pattern of phonetic accommodation, since the position of Difference #3 along the salience scale is not known. Instead, the pattern of accommodation within Difference #3 will be considered separately in §7.5.

6.2.4 Interspeaker Variation

In the preceding section, we saw that the dialectal differences varied in terms of how perceptually salient they were on average across all the participants. In addition, there is variation between the speakers in how perceptually salient these differences are for each of them. Table 17 shows the minimum and maximum accuracy rates across the 11 BAS participants, along with the mean and median, and Table 18 does the same for the 11 MS speakers. In all cases the means and medians are very similar, but we can see that across the participants there is a lot of variation in how perceptually salient these differences are.
This variation is likely due to various factors including each speaker’s perceptual sensitivity to acoustic detail (Namy et al. 2002) and level of experience with the opposing dialect. What this finding tells us is that speakers do not all perceive the dialectal differences in the same way; there is variation in how salient these differences are for the 22 speakers in terms of dialect recognition. The criteria-list approach to salience, discussed in §3, is unable to represent interspeaker variation like this.

6.3 The Pattern of Phonetic Accommodation

The main goal of this thesis is to determine what role perceptual salience plays in the pattern of phonetic accommodation during a single conversation between speakers of two different dialects of Spanish. In the previous section we discussed the results of the perception task designed to quantify how perceptually salient the differences between MS and BAS were for each of the participants who took part in the study. In the current section we address how these measures of perceptual salience are related to the pattern of phonetic accommodation found in the speech of the participants.
As discussed in §5.2, the participants read a word list containing lexical items that reflected each of the six dialectal differences in order to provide their baseline of pronunciation for these sounds. Next, each participant engaged in conversation with a speaker of the opposing dialect for approximately 40 minutes and then performed the word list again immediately after the conversation. To assess the pattern of accommodation, the baseline and post-conversation production were compared following the procedure explained in §5.3.2. As noted in that section, determining the pattern of accommodation requires consideration from two viewpoints: the size of the change from pre- to post-conversation, and the direction of the change. The change from pre- to post-conversation was calculated and standardized (following the procedure in §5.3.3) for each participant for each token. Changes towards the opposing dialect are considered convergence, while changes away from the opposing dialect are divergence.

While it is possible to represent both the magnitude and the direction of the change on a single value (0.8 is a convergence of 0.8, while -0.8 is a divergence of 0.8), the following analysis will consider the magnitude of the change separately from the direction of the change. This was necessary because if both the magnitude and direction are considered simultaneously, variation in the range of the magnitude of the changes between the dialectal differences is lost. For example, Figure 26 shows a series of kernel density plots\(^{19}\), which provide an illustration of the distribution of the value of the changes that were made, taking into account the direction (negative changes are divergences, positive changes are convergences), for the BAS speakers for each of the dialectal differences.

\(^{19}\) These kernel density plots were generated in R using the plot(density()) function in the *graphics* package (R Development Core Team 2012). Kernel density plots allow us to see how the data are distributed. The bandwidth value reflects the parameter used to smooth the data. The bandwidths given in Figure 26 and Figure 27 were generated automatically by R, although they can also be specified by the user. Higher bandwidths provide plots with greater smoothing, while lower values produce plots with less smoothing. The value of N gives the number of data points represented in the plot.
Figure 26: Distribution of change from pre- to post-conversation by BAS speakers by dialectal difference

The top left panel in Figure 26 shows the distribution of changes made on Difference #1 (BAS /s/ vs. MS /θ/). We can see that these changes range from large divergences (around -4)\(^{20}\) to large convergences (around 4), and that the distribution is centred close to 0. In contrast, the distribution of Difference #4 (BAS [x] vs. MS [χ]) is much narrower, ranging from -2 to 2, also centred about 0. The distribution of Difference #5 is slightly wider than that of Difference #4 and the distribution of Difference #6 is slightly wider still. This figure shows that while the changes made on all six differences are more or less centred on 0 (that is, the mean change is close to 0), there is real variation in how great of a change the BAS speakers make in either direction, both converging or diverging, that depends on the dialectal difference. If we take the magnitude and direction of the change into account at once, we lose this difference in the absolute value of the size of the change. The mean change made by BAS speakers on Difference #1 is 0.098, while the mean change for Difference #4 is 0.165. These means seem to suggest that a greater change was made on Difference #4 than on Difference #1; but, Figure 26 shows

\(^{20}\) Recall that the magnitude of the change was standardized following the procedure given in §5.3.3.
that clearly there were greater changes made (both towards and away from the MS group) on Difference #1 than there were on Difference #4. A similar situation is found within the MS speakers as shown in Figure 27.

![Distribution of change from pre- to post-conversation by MS speakers by dialectal difference](image)

Figure 27: Distribution of change from pre- to post-conversation by MS speakers by dialectal difference

The difficulty with using a value that reflects both the magnitude and direction of the change at once is that it cancels out large changes in two directions. The purpose of the current study is to investigate the effect of perceptual salience on the pattern of phonetic accommodation. If we consider all of the changes taking into account both the size and the direction at once and plot them against the perceptual salience of the dialectal differences (using the xylowess() function from the languageR package (Baayen 2011) to generate a scatterplot of the change by %Correct), we obtain the result in Figure 28. The line cutting through the data points is a lowess line, calculated using locally weighting scatterplot smoothing, and shows the trend throughout the data (Baayen 2008: 34). We can see that this line is essentially flat, showing no relationship between change and perceptual salience; however, we also notice that the range of changes (on the y-axis) gets larger as we move up in perceptual salience (along the x-axis). Given the plots
of the distributions of each dialectal difference in Figure 26 and Figure 27 this pattern is no surprise. Some of the differences make greater changes, whether they are convergences or divergences, than other dialectal differences. If we combine magnitude with direction, we lose this pattern.

Figure 28: *Plot of change (magnitude + direction) by perceptual salience including all speakers*

Furthermore, taking magnitude and direction together conflates two potentially very different processes. Laboratory studies of imitation have suggested that imitation may be a somewhat automatic process with speakers tending to imitate even when there was no obvious social motivation to do so, such as shadowing a pre-recorded model speaker (Goldinger 1998; Shockley et al. 2004; Delvaux & Soquet 2007; Babel et al. 2012). In contrast, Communication Accommodation Theory (CAT: Giles 1973) suggests that speakers use convergence and divergence strategically to arrive at the desired social distance between themselves and their interlocutors. On the one hand then, imitating may be an automatic process, but on the other hand, the direction of the change may depend on social factors. If we conflate the size of the change with the direction, we lose the opportunity to consider both aspects of the process: as automatic or as socially-motivated. It is for all these reasons that the magnitude of the change from pre- to post-conversation will be considered separately from the direction of the change.
Other studies have also investigated the magnitude and direction of accommodation separately, such as Kim, Horton & Bradlow (2011) and Black (2012).

Returning to the purpose of this section, in addition to the effect of perceptual salience there are various other factors that, based on previous findings in the literature, we might expect to affect the pattern of phonetic accommodation. These factors include dialect and gender of the participants (Namy et al. 2002; Babel 2009, 2011, 2012; Black 2012; Evans et al. in prep), the lexical frequency of the tokens that are produced (Goldinger 1998; Babel 2010; Nielsen 2011), and word-specificity (Goldinger 1998; Babel 2009; Nielsen 2011; Mielke et al. 2012). Word-specificity refers to an effect where the participants show greater imitation of words that they were explicitly exposed to as compared to other words that share a particular feature, but were not present in the exposure phase. In the current study this distinction corresponds to whether or not a particular token was included in the map task as one of the landmark labels. If a token was included in the map task labels, then both participants will have produced it. Other tokens not part of the map task will likely not have been produced.

Another factor that could potentially affect the pattern of accommodation concerns how much the distribution of a participant’s production of a particular dialectal variant overlaps with the opposing dialect’s production of the variant. The direction of the change (convergence or divergence) was defined relative to the mean of the opposing dialect, but how much two distributions overlap is a different measure. Babel (2009) concluded that speakers stay within their existing phonetic repertoire when converging towards a model speaker or conversational partner. This would seem to predict that the more a speaker has in common with another speaker (or dialect), the more we might expect him to be able to converge. To test this, a measure of overlap was included, calculated using linear discriminant analysis.

This measure of overlap constitutes another potential way to measure the degree of change from pre- to post-conversation, in addition to the method used here of comparing the mean of each dialectal difference in the two production sessions. We might expect that if a speaker’s mean production of Difference #1 (reflected in relative intensity) moves towards the mean of the opposing dialect, then this speaker’s degree of overlap with the opposing dialect’s pre-conversation production would likely also increase. Likewise, if a speaker’s mean moves

---

21 Although only one participant was the instruction “giver” at any given time, the participants were instructed to repeat their partner’s instructions, as discussed in §5.2.4.1.
away from the opposing dialect then their degree of overlap would likely also decrease. This is essentially what we see in the current data, as determined via a Pearson correlation of the ‘change’ in mean and the ‘change’ in overlap. The Pearson product-moment correlation coefficient provides a measure of linear dependence between two variables. This calculation gives the statistic $r$ which falls between -1 and 1. If $r = 0$ then there is no correlation between the two variables. An $r$ statistic of 1 indicates that the two variables are perfectly positively correlated, while a value of -1 means that the two variables are perfectly negatively correlated. The correlation between ‘change’ in means and ‘change’ in overlap was found to be positive and significant ($r = 0.318$, df = 130, $p = 0.0002$), indicating that as the change in means increases, so does the change in overlap. Figure 29 shows the distribution of the BAS speaker in Pair 6’s production of Difference #1 as compared to the distribution of the MS group’s production of this difference.²²

---

²² The BAS speaker in Pair 6 is compared to the entire MS group (rather than just to the MS speaker in Pair 6) because, as discussed in §5.3.2, the direction of accommodation (convergence or divergence) was defined for each speaker relative to the opposing dialect as a whole, rather than on an individual basis.

---

Figure 29: Amount of overlap in pre-conversation (left panel) and post-conversation (right panel) between BAS speaker in Pair 6 and entire MS group for Difference #1
convergence towards the MS group. In fact, the overlap in the pre-conversation session was 25.8%, while by the post-conversation session the overlap was 35.5%. This speaker’s standardized change in mean across all tokens was 1.06, also a convergence, which aligns with the change in overlap.

In contrast, Figure 30 shows an example of a decrease in overlap from pre- to post-conversation. The BAS speaker in Pair 4’s production of Difference #3 overlaps with the MS group by 21.1% in the pre-conversation session, but this overlap has dropped to only 5.3% by the post-conversation session, showing a divergence. This divergence is also reflected in the speaker’s (standardized) change in mean across all tokens as well, which was -0.78.

![Box plots showing overlap between BAS speaker and MS group for Difference #3](image)

**Figure 30:** Amount of overlap in pre-conversation (left panel) and post-conversation (right panel) between BAS speaker in Pair 4 and entire MS group for Difference #3

So we have seen that the change in means and the change in overlap are related, which makes sense. As discussed above, in the current study, the measure of accommodation is calculated in terms of how the mean COG changes from pre- to post-conversation. This measure allows a token-by-token indication of change since a particular acoustic measurement can be taken on an individual token, while to calculate ‘overlap’ we need two groups of data points. The results of the study will be discussed in terms of the magnitude of the change from pre- to post-conversation in §6.3.1 and in terms of the direction of the change in §6.3.2.
6.3.1 The Magnitude of Change

Before considering the effect of perceptual salience on the magnitude of the change, we first explore the general pattern in terms of each of various factors such as dialect, gender, lexical frequency, word-specificity, and overlap (with the opposing dialect). In order to examine the pattern of accommodation with respect to the magnitude of the change, we considered the absolute value of the (standardized) change (AbsChange). As an absolute value, the magnitude of the change is lower-bounded by 0. A kernel density plot of AbsChange (showing how the values of AbsChange are distributed for all tokens produced by all 22 speakers and all 6 dialectal differences) is shown in Figure 31. We can see that the distribution is skewed to the right and does not seem particularly normal.

![Kernel density plot of AbsChange](image)

Figure 31: Distribution of AbsChange (the absolute value of the change from pre- to post-conversation) across all participants and all dialectal differences

A quantile-quantile plot (qq plot) in Figure 32 shows that, as expected, there is a significant departure from normality. The qq plot provides a visual representation of the normality of a distribution. In a qq plot, the sample quantiles of the data, which are the values of the data for which below these values particular percentages of the data would be cut off, from 0 to 100% are plotted against the theoretical quantiles, which are the values obtained from a perfectly normal distribution (mean = 0 and standard deviation = 1). If the sample quantiles (generated from the experimental data) are normally distributed they will form a line 45 degrees from the x-
axis when plotted against the theoretical quantiles (Howell 2002: 77). In the qq plot below we can see that there is a large deviation from a 45-degree line and, in addition, there appear to be several outliers at the upper end.

Figure 32: Quantile-quantile plot of AbsChange (the absolute value of the change from pre- to post-conversation) across all participants and all dialectal differences

To improve the normality of the distribution and avoid outliers, the AbsChange data were log-transformed, giving Log AbsChange, as suggested by Baayen (2008: 92). The density of Log AbsChange is shown in Figure 33 below. Although the distribution is still not perfectly normal (now somewhat skewed to the left), it is an improvement over the untransformed AbsChange, as evidence by the much more normal qq plot for logAbsChange given in Figure 34, which also shows that the potential outliers in the untransformed data are no longer a problem.
As a result, the log-transformed absolute change will be used as the dependent variable in the regression analyses presented in the following section. The overall findings given in bar plots in the remainder of this section will be presented graphically using the non-transformed AbsChange, since this number is somewhat more intuitive. In the scatterplots, the log-
transformed value will be used because the relationships are more difficult to see when the x-axis is lower-bounded by 0 as in the non-transformed AbsChange.

The most basic finding is illustrated in Figure 35, showing that the mean absolute value of the standardized change (non-transformed) from pre- to post-conversation is almost identical for the MS group (mean AbsChange: 0.733) and the BAS group (mean AbsChange: 0.769), indicating that with all five dialectal differences (excluding Difference #3) and all participants taken together, the two dialects make virtually equal-sized changes from pre- to post-conversation. To verify that there is no difference by dialect, this factor (DIALECT) will be included in the linear regression in the next section.

![Figure 35: Mean standardized absolute change on all dialectal differences (excluding Difference #3) by dialect](image)

If we split this finding out by gender, we can see in Figure 36 that in both the MS and the BAS group, the female participants make a greater absolute change than their male counterparts. Previous studies have been inconclusive about whether males or females can be expected to accommodate more, but the tendency in Figure 36 suggests that females may make greater changes than males in this study. Gender will be entered into a regression model discussed in §6.3.1.1 to determine if it is a significant predictor of the magnitude of change.
Previous results have suggested that lower frequency tokens might be imitated (or converged to) more than the higher frequency tokens, as was found in Goldinger (1998), Babel (2010), and Nielsen (2011). Lexical frequency here was determined from The Corpus del Español (Davies 2002), which contains 100 million words. The frequencies are given as a log-transformed count of the number of times a particular lexical item appears in the corpus. Figure 37 below shows the relationship between the log-transformed lexical frequencies of the tokens included in the pre- and post-conversation word reading list and the log-transformed absolute value of the changes that the participants made.
Figure 37: Scatterplot of the log-transformed absolute value of the change and log lexical frequency for all participants, all dialectal differences (excluding Difference #3), and all tokens

The plot in Figure 37 does not suggest a strong relationship between these measures, but log lexical frequency will be included in the model described in the following section to determine what its effect on the magnitude of the change is.

Babel (2009) concluded that speakers stay within their existing phonetic repertoire when converging towards a model speaker or conversational partner. This would seem to predict that the more a speaker has in common with another speaker, the more we might expect him to be able to converge. To test this prediction, I measured the amount of overlap between the distributions of each speaker and the opposing dialect in terms of the acoustic measures for each of the dialectal differences. The amount of overlap was quantified using linear discriminant analysis. For example, to quantify the overlap of the BAS speaker in Pair 3’s production of /x/ with the entire MS group’s production of that sound I conducted a linear discriminant analysis including all of the BAS speaker’s measurements for /x/ and all of the measurements for all of the MS speakers, using dialect as the grouping variable, and Onset F2 (the acoustic correlate used to characterize Difference #4) as the independent variable. The proportion of the BAS speaker’s tokens that would be incorrectly classified as belonging to the MS group based on F2 onset represents the amount of overlap between the two distributions. The greater the percentage that can be correctly classified, the less the overlap between the distributions. The chart below in
Figure 38 shows that there is a negative relationship between overlap and the log-transformed AbsChange: the less overlap there is between the variable and the opposing dialect the larger the change from pre- to post-conversation. Overlap will be entered into the regression discussed in the following section and considered more carefully in §7.3.

Several imitation studies have found an effect of word-specificity, that is, tokens that participants were exposed to showed more imitation than those they were not exposed to. In the current study, a subset of the words that the participants read in the pre- and post-conversation word lists were also included as names in the landmarks on the maps used in the map task. As such, the participants would have been exposed to their conversational partner producing the words in the landmark names, while they likely did not hear their partner say the words that were not part of the map task. The prediction from the results of studies investigating word-specificity is that the participants would converge more on the tokens that were part of the map task, than those that were not.
The graph in Figure 39 does not seem to suggest that such an effect will be found here since the size of the change from pre- to post-conversation appears almost identical whether the token was included in the map task or not, for both dialects, but to be sure, this factor (LANDMARK TOKEN) will be included in the regression in the next section.

Although the current section is focused on the magnitude of the effect, we can incorporate the direction of the change here by comparing the magnitude of the changes made on tokens that were converged upon and tokens that were diverged upon.
Figure 40 shows that for both the BAS and the MS speakers, the tokens that were converged upon showed a greater change than those that were diverged upon. This factor (DIRECTION) will be included in the regression in the next section.

So overall we have examined the magnitude of the change from pre- to post-conversation by dialect, gender, lexical frequency, overlap, word-specificity, and direction. We turn now to the effect of perceptual salience on the magnitude of the change.

6.3.1.1 Perceptual Salience and the Magnitude of Change

As discussed in previous chapters, the literature makes differing predictions regarding the expected effect of perceptual salience on the pattern of phonetic accommodation. Trudgill (1986: 11) predicts that speakers will converge towards their conversational partners more on the dialectal differences that are more salient to them than on those that are less salient. In contrast, recent experimental findings from Babel (2010) and Kim et al. (2011) suggest that it is the least salient dialectal differences that show the most convergence. Evans et al (in prep) found that it was the most salient dialectal differences that speakers diverged on more. To find support for Trudgill’s position we would expect that the participants in the current study would converge more often on the more salient differences and that the changes made on these
differences would be greater than those made on the less salient differences. Support for the findings of Babel (2010) and Kim et al. (2011) would entail the participants converging more often on the least salient differences and making greater changes on these differences as compared to the more salient differences. If the results of the current study support the findings of Evans et al. (in prep), we would expect that the participants would diverge more often on the most salient dialectal differences and making a greater change on these differences as compared to the less salient differences.

To get an idea of the pattern of the magnitude of the change, we can again use the xylowess() function from the languageR package (Baayen 2011) to generate a scatterplot of the log-transformed AbsChange data (the absolute value of the changes made from pre- to post-conversation for each token and each participant) by dialect which includes a trend line in the data. We can see in Figure 41 that there appears to be a positive relationship between perceptual salience and the magnitude of the change for both dialects. As perceptual salience of the dialectal differences (excluding Difference #3) increases, the participants make greater changes on those dialectal differences.

![Figure 41: Scatterplot of the log-transformed absolute value of the change perceptual salience (%-Correct) for all participants and all dialectal differences (excluding Difference #3) by dialect](image)

To see if the relationship between the magnitude of the change and perceptual salience is statistically significant (along with the other predictor variables discussed above), a linear
The mixed-effects model was built with the log-transformed absolute value of the change from pre- to post-conversation (LOG ABSCHANGE) for each token produced by each of the 22 participants in the study as the dependant variable.

The fixed effects in the model were %-CORRECT of the dialectal difference contained in the token (with the mean subtracted to centre it about 0), DIALECT and GENDER of the participant, the log-transformed LEXICAL FREQUENCY of the token, whether or not the token was included as a LANDMARK NAME in the map task, the OVERLAP between the distribution of the variant and the opposing dialect, and the DIRECTION of the change (convergence or divergence). The random effects were token and subject and the model included a by-subject random slope for %-CORRECT. A likelihood ratio test determined that the more complex model incorporating the by-subject slopes for %-CORRECT was justified ($\chi^2=30.087$, p<0.0001, df=2). As in the model described in §6.2.2, the categorical predictor variables (DIALECT, GENDER, LANDMARK NAME, and DIRECTION) were sum-coded and centred about 0 (for DIALECT: dialect_BAS = -0.5, dialect_MS = 0.5; for GENDER: gender_Female = -0.5, gender_Male = 0.5; for LANDMARK NAME: landmarkname_NO = -0.5, landmarkname_YES = 0.5; for DIRECTION: direction_DIV = -0.5, direction_CON = 0.5). The reference levels were BAS for DIALECT, Female for GENDER, NO for LANDMARK NAME, and DIV(ergence) for DIRECTION.

Table 19 below shows the output of the model, including only the terms that reached significance. As expected from the plot in Figure 41, %-CORRECT was significant with a positive coefficient, indicating that there is a significant positive relationship between salience and change: as salience increases, the participants make greater changes from pre- to post-conversation.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coef. Est.</th>
<th>StdErr</th>
<th>t-value</th>
<th>p-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.91792</td>
<td>0.06357</td>
<td>-14.441</td>
<td>&lt;0.0001</td>
<td>***</td>
</tr>
<tr>
<td>%-CORRECT</td>
<td>0.80277</td>
<td>0.17837</td>
<td>4.501</td>
<td>&lt;0.0001</td>
<td>***</td>
</tr>
<tr>
<td>OVERLAP</td>
<td>-0.59112</td>
<td>0.18414</td>
<td>-3.210</td>
<td>0.001</td>
<td>**</td>
</tr>
<tr>
<td>DIRECTION_DIV</td>
<td>-0.27429</td>
<td>0.04319</td>
<td>-6.351</td>
<td>&lt;0.0001</td>
<td>***</td>
</tr>
</tbody>
</table>

This significant main effect of %-CORRECT indicates that as a group, the participants make greater changes on the more perceptually salient dialectal differences. We can delve deeper into this finding by examining the relationship between the magnitude of the change and perceptual...
salience at the individual speaker level using the by-subject coefficients for %-CORRECT, given in Table 20. The by-subject coefficients are the result of including a by-subject slope for %-CORRECT as a random effect in the model. These values reflect the deviation of each speaker from the coefficient generated for the entire group, both in terms of the direction and the extent of the deviation (Drager & Hay 2012). That is, the by-subject coefficients for %-CORRECT show how perceptual salience affects the magnitude of the changes made by each individual speaker. As shown in Table 20, the by-subject coefficient is positive for every single participant, showing that there is a positive relationship between salience and the magnitude of the changes made for each speaker.

<table>
<thead>
<tr>
<th>Pair</th>
<th>BAS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.085</td>
<td>1.475</td>
</tr>
<tr>
<td>2</td>
<td>0.312</td>
<td>0.566</td>
</tr>
<tr>
<td>3</td>
<td>0.519</td>
<td>1.239</td>
</tr>
<tr>
<td>4</td>
<td>0.939</td>
<td>1.227</td>
</tr>
<tr>
<td>5</td>
<td>0.257</td>
<td>0.422</td>
</tr>
<tr>
<td>6</td>
<td>1.005</td>
<td>0.573</td>
</tr>
<tr>
<td>7</td>
<td>0.784</td>
<td>0.715</td>
</tr>
<tr>
<td>8</td>
<td>1.764</td>
<td>1.297</td>
</tr>
<tr>
<td>9</td>
<td>0.279</td>
<td>0.696</td>
</tr>
<tr>
<td>10</td>
<td>1.115</td>
<td>0.268</td>
</tr>
<tr>
<td>11</td>
<td>0.972</td>
<td>0.140</td>
</tr>
</tbody>
</table>

The slopes vary in how steep they are indicating that the participants vary in how sensitive they are to salience, but all are positive. For example, the MS speaker in Pair 1 has a relatively large slope of 1.475, meaning that as salience increases she makes a much larger change. The MS speaker in Pair 11, on the other hand, is much less affected by salience since her slope for %-correct is 0.140, indicating that she differs little from the trend of the overall model. Nevertheless, all of the speakers make greater changes as salience increases. The significance of this finding in terms of the mechanism behind accommodation will be discussed in §7.7.

OVERLAP was significant and the coefficient is negative, meaning that as overlap between the distribution of a particular variant and the distribution of the opposing dialect
increases, the magnitude of the change from pre- to post-conversation decreases. This result will be discussed further in §7.4.

In addition, the DIRECTION term was highly significant and the coefficient was negative, which means that the size of the change made on the tokens that were diverged upon was significantly lower than the size of the change made on the tokens that were converged upon, as we might have expected given the graph in Figure 40. This will be considered in §7.8.

Since the graph in Figure 35 showed that the mean AbsChange made by the two dialects was almost identical, it is no surprise that the DIALECT term did not reach significance. Although the graph in Figure 36 showed that the female participants made greater changes than their male counterparts, the main effect of GENDER was also not significant. This could potentially be because there were only two female pairs (4 female participants) in the study as compared to nine male pairs (18 male participants).

As expected from the scatterplot in Figure 37, LEXICAL FREQUENCY was not found to significantly affect the magnitude of the change from pre- to post-conversation. The effect of LANDMARK NAME (word-specificity) was also found not to be significant. We might have expected that those words that were included in the map task and therefore were heard and spoken by the participants might show a greater change from pre- to post-conversation. Such a result would be in line with Goldinger (1998)’s finding that imitation of specific tokens increased when participants had heard more repetitions of those tokens in a pre-imitation listening phase. In contrast, Nielsen (2011) found no significant effect of previous exposure, so the expected effect of word-specificity is not completely clear.

Having explored the pattern of phonetic accommodation in terms of the magnitude of the change that the participants made from pre- to post-conversation for the five dialectal differences, we turn now to the direction of these changes, as either converging towards or diverging away from the opposing dialect, first considering the general pattern, and then the specific relationship between perceptual salience and the likelihood of converging (or diverging).
6.3.2 The Direction of Change

As discussed in §5.3.2, changes that moved towards the mean of the opposing dialect were coded as convergences, while changes that moved away from the mean of the opposing dialect were coded as divergences. Thus, when considering the direction of the change, the relevant variable is binary: a speaker either converged or diverged on a particular token. As a result, when examining the pattern of accommodation in terms of the direction of the change we are considering the proportion of tokens on which the speakers converged or diverged.

Figure 42 shows that both the BAS group and the MS group were more likely to converge on the tokens they produced than they were to diverge since each dialect converged on almost 60% of the tokens (diverging on about 40%).

![Proportion of tokens converged upon by dialect](image)

Figure 42: Proportion of tokens converged upon by dialect

Splitting this result out by gender, Figure 43 shows the proportion of tokens converged upon for each dialect by gender. We can see that in both dialects, the females converged more frequently than the males, although the difference between the genders is greater within the BAS group.
Considering the proportion of tokens converged upon by dialectal difference, Figure 44 shows that all of the dialectal differences were converged upon more often than they were diverged upon by both dialects. The horizontal line in Figure 44 marks the 50% point on the y-axis and we can see that all of the bars extend beyond this point. In addition, we can see that the MS speakers converge more often than the BAS speakers 3 of the 5 dialectal differences (excluding Differences #1 and #4).
As we saw in §6.3.1.1, there was no significant main effect of lexical frequency on the magnitude of the change from pre- to post-conversation. Figure 45 below suggests that there will be no significant effect of lexical frequency on the likelihood of converging on any given token either.

Figure 45: Scatterplot of the likelihood of converging (DIV=0, CON=1) by the log-transformed lexical frequency for all participants and all dialectal differences (excluding Difference #3)

In contrast, there may be an effect of overlap between a participant’s pre-conversation production of a dialectal variant and the opposing dialect’s production of that variant on the likelihood of converging, since Figure 46 shows a potentially significant positive relationship between overlap and proportion of tokens converged upon.
Lastly, Figure 47 below gives the proportion of tokens converged on by speakers of each dialect, split by whether or not the tokens were included as landmark names in the map task. We see that for both dialects the tokens that were included in the map task were slightly more likely to be converged upon than those that were not in the map task, although the size of the difference is small. This factor, LANDMARK NAME, is included in the model to determine its effect on the likelihood of converging.
In this section, we examined the direction of the change from pre- to post-conversation by dialect, gender, the individual five dialectal differences, lexical frequency, overlap, and word-specificity. We turn now to the effect of perceptual salience on the direction of the change.

6.3.2.1 Perceptual Salience and the Direction of Change

As we saw in Figure 44 above, there seems to be variation in the likelihood of converging that depends on the dialectal difference that the tokens reflect, as well as potentially an effect of the dialect of the participant, since we saw differences in the likelihood of converging between the dialects within particular dialectal differences. For example, we saw that the MS speakers were more likely to converge than the BAS speakers on 3 of the 5 dialectal differences, except Differences #1 and #4. In addition, both dialect groups were more likely to converge than diverge on all of the dialectal differences. To what extent can this variation in patterning be explained by the perceptual salience of the dialectal differences?

To visualize the relationship between the likelihood of converging and perceptual salience we again make use of the xylowess() function to generate a scatterplot of the data for each dialect. As the graph in Figure 48 shows, both dialects seem to show a quadratic relationship between perceptual salience and the likelihood of converging, where the dialectal differences at the two extremes of salience are less likely to be converged upon than those in the middle, although the pattern is much flatter in the case of the MS speakers.
To investigate the factors affecting the likelihood of convergence a logistic mixed effects regression model with a binomial distribution was built with DIRECTION of the change (convergence or divergence) as the dependent categorical variable. Since the plot in Figure 48 suggests a quadratic relationship between direction and salience, the model included $\%$-CORRECT$^2$ as a fixed effect. The other fixed effects were DIALECT and GENDER of the participant, the log-transformed LEXICAL FREQUENCY of the token, whether the token was included as a LANDMARK NAME or not, and the OVERLAP between the production of a particular dialectal difference by the speaker and the opposing dialect, as well as an interaction between $\%$-CORRECT$^2$ and DIALECT, since the plot in Figure 48 suggests that the relationship between salience and the likelihood of converging may differ between the dialects. Again, the reference levels were BAS for DIALECT, Female for GENDER, and NO for LANDMARK NAME. The random effects were subject and token, and a by-subject random slope for $\%$-CORRECT.

The output of the model is given in Table 21 below. The only significant terms (besides the intercept) were GENDER and OVERLAP.
Table 21: Model summary: linear mixed effects model with DIRECTION as dependant

Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coef. Est.</th>
<th>StdErr</th>
<th>z-value</th>
<th>p-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.48190</td>
<td>0.06142</td>
<td>7.846</td>
<td>&lt;0.0001</td>
<td>***</td>
</tr>
<tr>
<td>%CORRECT</td>
<td>-0.05683</td>
<td>0.22955</td>
<td>-0.248</td>
<td>0.804</td>
<td></td>
</tr>
<tr>
<td>GENDER_Male</td>
<td>-0.29341</td>
<td>0.12150</td>
<td>-2.415</td>
<td>0.016</td>
<td>*</td>
</tr>
<tr>
<td>OVERLAP</td>
<td>0.74571</td>
<td>0.30371</td>
<td>2.455</td>
<td>0.014</td>
<td>*</td>
</tr>
</tbody>
</table>

Although the plot in Figure 48 above suggested a quadratic relationship between salience and the likelihood of convergence, the %CORRECT^2 term was not found to be significant. And although a linear %CORRECT term was not expected to be significant from the plot in Figure 48, various runs of the model found that a by-subject random slope for %CORRECT did result in a model that was significantly different from one without such a term as determined via a likelihood ratio test ($\chi^2 = 9.7522$, $p = 0.002$, df = 1), but was not significantly different from a model with a by-subject random slope for %CORRECT^2. As such, the by-subject random slope for %CORRECT was included in the model.

Examining the by-subject coefficients for %CORRECT, given in Table 22, sheds light on the relationship between perceptual salience and the likelihood of converging at the individual speaker level. As discussed in §6.3.1.1, the by-subject coefficients show us how individual speakers deviate from the pattern at the group level. In this case, positive by-subject coefficients indicate that the individual speakers are more likely to converge as salience increases, while negative coefficients indicate that the individual speakers are less likely to converge as salience increases. The coefficients in Table 22 show a lot of variation both in terms of the size of the coefficient per participant as well as the direction of the slope (positive or negative). This means that the 22 participants vary in how perceptual salience affects their likelihood of converging on any given token. Some of the participants, such as the MS speaker in Pair 6, with a large positive slope (1.011), are much more likely to converge as perceptual salience increases, while others such as the BAS speaker in Pair 2, with a relatively large negative slope, are much less likely to converge as perceptual salience increases. Other participants show varying degrees of sensitivity to perceptual salience, such as the MS speaker in Pair 2, whose random slope is 0.048, indicating that his likelihood of converging is almost not affected by perceptual salience at all.
Table 22: By-subject coefficients for %-CORRECT

<table>
<thead>
<tr>
<th>Pair</th>
<th>BAS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.072</td>
<td>-0.450</td>
</tr>
<tr>
<td>2</td>
<td>-0.976</td>
<td>0.048</td>
</tr>
<tr>
<td>3</td>
<td>0.328</td>
<td>0.196</td>
</tr>
<tr>
<td>4</td>
<td>0.734</td>
<td>-0.105</td>
</tr>
<tr>
<td>5</td>
<td>0.403</td>
<td>0.088</td>
</tr>
<tr>
<td>6</td>
<td>-0.140</td>
<td>1.011</td>
</tr>
<tr>
<td>7</td>
<td>0.189</td>
<td>-0.699</td>
</tr>
<tr>
<td>8</td>
<td>-0.692</td>
<td>-0.109</td>
</tr>
<tr>
<td>9</td>
<td>-0.366</td>
<td>-0.306</td>
</tr>
<tr>
<td>10</td>
<td>0.299</td>
<td>0.732</td>
</tr>
<tr>
<td>11</td>
<td>-0.688</td>
<td>-0.826</td>
</tr>
</tbody>
</table>

That the by-subject coefficients for %-CORRECT are so varied between the participants may explain why there is no significant main effect of perceptual salience: how perceptual salience affects each participant is different. As salience increases, some are more likely to converge and others are less likely to converge, and the extent of this variation is beyond what we might expect from random variation. With so much interspeaker variation, it is no surprise that the main effect for salience was not significant. In general, it may also be no surprise that there is no consistent effect between participants of perceptual salience on the likelihood of converging. Under Communication Accommodation Theory (Giles 1973), convergence and divergence are used strategically by speakers to arrive at the desired social distance between themselves and their interlocutor. Whether a particular speaker will want to converge or diverge (and by how much) is a very personal characteristic, and not something that would be expected to be the same for every participant in an experiment. This idea will be explored further in §7.1.

Returning to the model given in Table 21, the negative coefficient of the significant GENDER term indicates that the male participants were less likely to converge than the females. The mean proportion of convergence was 57.8% for the males and 63.2% for the females for a difference of 5.4% between the genders. The OVERLAP term was significant with a positive coefficient, indicating that as overlap between a speaker’s production of a variant and the opposing dialect’s production of that variant increases, the likelihood of converging increases as well. This result will be discussed further in conjunction with the significant finding for OVERLAP in terms of the magnitude of the effect in §7.3.
As expected from the plots given in the previous section, there was no significant effect of DIALECT, LEXICAL FREQUENCY or LANDMARK NAME.

6.4 Summary of Results

The preceding sections examined how the perceptual salience of the six dialectal differences between BAS and MS affected the pattern of phonetic accommodation both in terms of the magnitude of the change from pre- to post-conversation and the direction of the change, that is, the likelihood of converging or diverging. The results showed that there was a significant effect of perceptual salience on the magnitude of the change for both dialects: as perceptual salience increased, the participants made greater changes from pre- to post-conversation. This effect held for both dialects and was consistent even at the individual speaker level. On the other hand, there was no significant main effect of perceptual salience on the likelihood of converging or diverging. However, including a by-speaker random slope for %-correct resulted in a model that accounted for more of the variance in the likelihood of converging than a model without such a term, suggesting that perceptual salience affected the likelihood of converging significantly, but differently between participants. An examination of the by-subject coefficients for salience confirmed this, finding that some of the participants were more likely to converge on the more salient differences, others were less likely to converge on the more salient differences, and some showed very little effect of salience on the likelihood of converging. The following chapter attempts to explain these results.
Chapter 7
Discussion

The purpose of this thesis was to investigate the pattern of phonetic accommodation by speakers of two dialects of Spanish, with a particular focus on how the perceptual salience of the various differences between the dialects might affect the pattern. The question of how salience interacts with accommodation is an interesting one because very few studies have considered it previously, but those that have touched on this question have found completely opposing results. In earlier chapters, I suggested that part of the reason for differing results between the findings of previous studies could stem from variation in how perceptual salience had been defined. In §3, I reviewed the two main ways of defining salience that we find in the literature: the criteria list approach and the approach that salience is a context-dependant, measurable property of linguistic variables. In that chapter, I argued that which differences between two dialects will be considered perceptually salient to speakers of a given dialect depends crucially on the context in which those sounds are perceived and what social meaning is attributed to the use of a particular variant. As discussed, other researchers have used a similar methodology to that incorporated in this study to determine how perceptually salient particular variables are for identification of the social characteristics of the speakers who use those variables, such as regional dialect and ethnicity (Graff, Labov & Harris 1986; Fridland et al. 2004; Thomas & Reaser 2004; Torbert 2004, 2010).

Using this methodology I was able to quantify how perceptually salient the dialectal differences under investigation between BAS and MS are for each of the 22 participants who took part in the experiment. Crucially, the perceptual salience of these differences is measured in terms of how accurately the participants were able to identify their own dialect on the basis of the dialect variants. In this way, perceptual salience here is defined within the context of dialect recognition. The variables whose presence facilitates dialect recognition are therefore considered more salient for dialect recognition. Conversely, those variables that do not increase the accuracy with which the participants can identify their dialect are considered less salient for dialect recognition. This finding says nothing about how perceptually salient those same variables would be for identification of other speaker characteristics such as ethnicity, socio-economic class, age or gender, or about how great the objective phonetic distance is between the variants of the two dialects.
With this measure of salience in hand I was able to investigate how these individual-specific values interacted with the pattern of accommodation that took place as a result of a conversation between a BAS speaker and an MS speaker. The results of the study, as discussed in §6, found that there is a significant effect of perceptual salience on the magnitude of the changes that speakers make in their production from pre- to post-conversation in that as perceptual salience of the dialectal differences increased, the participants made a greater change. In contrast, there seemed to be an effect of perceptual salience for the direction of the change (converging or diverging away from the opposing dialect norms) at the level of the individual, but the effect was not consistent at the group level. I attempt to explain these findings by appealing to social motivations in §7.1, phonological contrast in §7.2, overlap in §7.3, and gender in §7.4.

In §7.5, I examine the pattern of accommodation found for Difference #3 and argue that this difference is likely much more perceptually salient than the results of the perception task suggested.

As we have seen, there was much inter- and intra-speaker variation in the current study. In §7.6, I will discuss individual-level factors that may have contributed to this variation. After that, I will examine how the results of this study weigh in on the debate regarding the mechanism behind accommodation in §7.7. In §7.8, I explore convergence and divergence in terms of whether these adjustments are really mirror images of each other. In §7.9, I discuss how the effect of perceptual salience might be expected to extend beyond phonetics to other domains of speech and explore this idea with respect to the conversations of 3 of the BAS participants. Lastly, I will consider the predictions that the findings of my study make for longer-term patterns of accommodation such as second-dialect acquisition as well as community-level changes in dialect contact situations and sound change in general in §7.10.

7.1 Social Motivations

The results of this study showed that perceptual salience affected the magnitude of the changes made from pre- to post-conversation since the participants made greater changes as perceptual salience increased, but that there was no consistent effect of salience on the direction of the change. In this section, I propose that both of these findings can be partly explained by
appealing to the social motivations of the individual speakers, which are expressed well using Communication Accommodation Theory (CAT: Giles 1973).

As previously discussed, CAT is a social-psychological model of changes in speech style during interaction. CAT posits that speakers use convergence and divergence strategically to attain the desired social distance between themselves and their interlocutors with the main motivation for these adjustments in speech being social in that speakers have a “need (often unconscious) for social integration or identification with another” (Giles et al. 1991: 18). Speakers are predicted to converge in order to decrease social distance between themselves and their interlocutors, to increase intelligibility, or to improve communication, or to diverge in order to increase social distance, accentuate distinctiveness, show disdain, or signal group membership (Shepard, Giles & Le Poire 2001). If convergence and divergence are used strategically by speakers to minimize social distance or to accentuate distinctiveness, respectively, then we might expect that speakers who were inclined to want to do these things would focus their efforts on the linguistic variables that would be most likely to have the greatest impact in achieving their goals: the most salient linguistic variables. For those who wish to minimize social distance and converge towards the opposing dialect, doing so on the most salient differences would likely make the greatest impact. Similarly, those who wish to accentuate the distinctiveness between their native dialect and the opposing dialect, diverging on those most salient differences would again have the greatest impact. This may partially explain why we see large changes in both directions occurring on the two most salient dialectal differences: as the perceptual salience of the dialectal differences increases, so does the potential for making an impact in terms of perceived similarity (or dissimilarity) to the opposing dialect.

The social motivation of the individual speaker then may be indirectly reflected in the perceptual salience of the dialectal difference. Since perceptual salience here is defined within the context of dialect recognition, the more salient differences are those that reflect a more reliable marker of (at least one of) the dialects. The more salient differences are more meaningful as markers of the dialects, so the speakers in the study use these more salient differences for whichever purpose they have: converging or diverging. Whether a speaker wants to converge or diverge, he can make the greatest impact in doing so by making large changes on the most salient differences.
As we have seen, perceptual salience does not play the same role for all speakers as to whether they will converge or diverge. As suggested earlier, this is perhaps not unexpected since whether a speaker will be inclined to want to shift towards or away from another speaker or another dialect is a very personal characteristic of individuals, which will depend on many factors such as the strength of identification with the D1 and D2 (Siegel 2010: 106). In this way, social motivations affect both the magnitude and the direction of the change.

7.2 Phonological Contrast

If we examine the likelihood of converging within the two most salient dialectal differences, we find an asymmetry between the dialects. The BAS speakers converged more often on Difference #1 (60% of tokens) than they did on Difference #2 (54%), while the MS speakers converged more on Difference #2 (57%) than they did on Difference #1 (54%). In the case of the BAS speakers, the difference in likelihood of converging between the two most salient difference was marginally significant ($z(707) = -1.81$, $p = 0.07$), while the difference was not significant for the MS speakers. These two dialectal differences did not differ from each other in terms of how perceptually salient they were, so we cannot appeal to salience to explain this asymmetry. What can account for the difference?

One possibility relates to the status of the variants with respect to phonological contrast. A significant distinction between these two dialectal differences from the perspective of the MS speakers is that Difference #1 reflects a phonemic contrast of MS, whereas Difference #2 does not. The two variants of Difference #1 are /s/ and /θ/, both of which are phonemes of MS. MS speakers need to maintain this contrast, which has a relatively heavy functional load being involved in many minimal pairs, such as *casa* /ka.sas/ ‘house’ and *caza* /ka.θas/ ‘hunt.3sg’. On the other hand, Difference #2 does not reflect a phonemic contrast in MS, since the BAS variant /ʃ/ is not found in the segmental inventory of MS. As a result, for the MS speakers, converging towards the BAS variant of Difference #2 would not threaten a native contrast in the way that converging towards the BAS variant of Difference #1 could. As discussed in §3, the status of a pair of dialect variants in phonological contrast was predicted by Trudgill (1986) and others to be a predictor of the salience of the difference. It is likely the case that Difference #1 reflecting a phonological contrast in MS contributed to the finding that this difference is highly salient for
the MS speakers as well as having a mitigating effect on whether MS speakers will converge on it.

A similar result is found for Difference #2, which represents a phonemic contrast in BAS, but not in MS. In BAS, /ʃ/ contrasts with the palatal glide [j] (an allophone of the high front vowel /i/ - Quilis 1993; Hualde 1991, 1999, 2005) in minimal pairs such as yerba [ʃeɾ.βa] ‘maté’ and hierba [jeɾ.βa] ‘grass’. In addition, [ʃ] alternates with [j], where the glide occurs in coda position as in buey [bwej] ‘ox’, and the fricative occurs in onset position, as in bueyes ['bwe.ʃes] ‘oxen’. The glide also appears when part of a complex onset, as in siendo ['sjen.do] ‘being’. As noted above, the BAS speakers were marginally more likely to converge on Difference #1 (which does not reflect a contrast in their dialect) than they were on Difference #2 (which does). In addition, the MS speakers were more likely to converge on Difference #2 than the BAS speakers (although the difference in probability of convergence was not significant). In essence, the speakers for whom a contrast would be threatened by converging do so less than those for whom no contrast is threatened. This asymmetry in the contrastive status of the two most salient dialectal differences might help explain the difference in patterning between the two dialects.

Nielsen (2011) also found that a linguistic factor such as phonological contrast could constrain the pattern of phonetic accommodation. In that study, participants imitated voiceless stops with extended VOT, but not those with reduced VOT. Nielsen concluded that the extended VOT did not threaten any contrast, while the reduced VOT started to encroach on the values expected for voiced stops in English. We will see further evidence that phonological contrast plays a role in the pattern of phonetic accommodation when we consider the results for Dialectal Difference #3 (BAS [s] versus MS [ʃ]) in §7.5.

7.3 Overlap

In §6, it was found that the degree of overlap, calculated using linear discriminant analysis, between a participant’s production of a given dialect variant and the opposing dialect’s production of that variant affected both the magnitude of the change as well as the direction of the change. As overlap increased, the magnitude of the change decreased, while the likelihood of convergence increased.
The motivation for considering overlap came from the results of Babel (2009). In that study, Babel suggested that participants use variants already present in their phonetic repertoire when converging towards another speaker, noting that “Participants are selecting from pre-existing variants in order to approximate the auditory targets of the model talker, but they are not encoding new speech production targets simply for this task” (Babel 2009: 129). In Babel’s study, the overlap between the participants and the model talker (in the baseline production phase) was not investigated, instead she argued that because the low vowels were more variable in terms of formant values (due to a greater effect of pitch accent on low vowels than high vowels), the participants were able to make greater changes on the low vowels. However, a participant’s production of a particular sound could be highly variable, but still not overlap at all with his interlocutor’s. If participants do not generate new production targets when converging, we might expect that a participant whose production overlaps more with his interlocutor would have more opportunities to converge since the targets needed are already present in his repertoire. In fact, this is what was found in §6.3.2.1, where the logistic regression model found a significant positive relationship between overlap and the likelihood of converging. As overlap increased, the likelihood of converging increased. This result provides support for Babel’s suggestion that participants do not encode new production targets in order to converge.

However, Babel also found that her participants made greater sized convergences on the low vowels which were more variable than high vowels, and argues that greater variability predicts larger convergences. The more variable a speaker’s production is, the more potential there is for overlap with another speaker or dialect (although the former does not necessarily imply the latter). If greater variability predicts greater overlap (at least to some degree), then we might expect that greater overlap would also predict larger convergences. However, as we saw in §6, as overlap between a participant’s production and the opposing dialect’s production of dialectal difference increased, the magnitude of the change made from pre- to post-conversation decreased. The scatterplot illustrating this relationship is given in Figure 49.
Figure 49: Scatterplot of the log-transformed absolute value of the change from pre- to post-conversation by %-overlap for all participants and all dialectal differences (excluding Difference #3)

The result is a negative relationship, one that is reflected in the significant negative coefficient in the model of the magnitude of the change presented in §6.3.1.1. This finding holds if we examine only the tokens that were converged on, since Babel’s finding really only concerns the magnitude of convergences and not divergences, as shown in Figure 50.

Figure 50: Scatterplot of the log-transformed absolute value of the change from pre- to post-conversation by %-overlap for all participants and all dialectal differences for all tokens converged upon (excluding Difference #3)
If the participants in Babe (2009)’s study converge more on the low vowels than the high vowels as a result of them being more variable, then greater variability does not predict greater overlap with another speaker or dialect, since, as found in this study, as overlap increased, the magnitude of convergence decreased.

As noted above, Babel (2009) suggests that speakers stay within their phonetic repertoire when converging towards another speaker, but how can we really know what is in a speaker’s inventory? How can we know which variants are available to be produced? If we assume an exemplar-style model (Johnson 1997; Pierrehumbert 2001) then, in principle, everything that the speaker has perceived is available to be produced, but the likelihood that she will produce a given variant depends on the activation level of that variant. If recent exposure to an uncommon variant raises its activation level, then the likelihood of choosing that variant for production increases. In this way, using the pre-conversation production of a particular linguistic variable as a measure of the speaker’s pool of production variants is only reflective of the speaker’s general way of speaking and her recent interactions in that it tells us which variants have high activation levels, but it tells us nothing about which variants might have low activation levels, causing them not to have been produced.

Following an exemplar model to the letter, I assume that all of the participants in the current study have the same distribution of perception variants because they have likely all heard all the sounds from both dialects at some point in their lives. Within a speaker, the variants from the two dialects will differ in terms of activation levels as a reflection of the much more common experience with their own dialect, and there may be some distortion due to incorrect weighting that stems from inaccurate stereotypes about the opposing dialect, but I assume that they all start with the same pool of perception (and thus, production) variants. Under an exemplar approach then, they all have the potential to converge 100% towards each other. Relating this to the idea of staying within a repertoire, we have no real need to posit that speakers are encoding new production targets, since we assume they have all of them already. So, exposure to those other-dialect variants would be predicted to raise the activation of those less common variants, potentially enough to trigger their selection for production (giving rise to a convergence). To the extent that this describes what happens during perception and production, a speaker is staying within their phonetic repertoire since they are using previously-encoded variants, even if those variants had thus-far only been used perceptually.
This approach makes sense in light of the finding that there is a negative relationship between overlap and the magnitude of change as well as the positive relationship between overlap and the likelihood of convergence. If a participant has the same pool of variants at their disposal as the speakers of the opposing dialect, then when there is less overlap between the distributions of a particular variable pre-conversation, there is a greater chance that during the conversation the participant will hear his opposite-dialect partner using variants that have a low activation level in the participant’s pool of variants. If he hears these less common variants, their activation will increase and this could cause him to choose these variants for production, generating a convergence towards that mean of the opposing dialect. Since the variant is not part of his usual way of speaking, in that it was not part of his pre-conversation production, producing such a variant would cause a large shift in his mean production. Such a change would only happen when the low-activation variant reaches a certain degree of activation, thus, potentially most of the time, that degree of activation would not be reached and the speaker would instead choose an exemplar with a high activation: one within his normal range of production.

In this way, when overlap is low, participants are more likely to hear variants that if they imitated them, would cause a large change, but which are not as likely to be imitated since these variants would require being heard a certain number of times before an adequate activation was reached for them to be chosen for production. On the other hand, when there is more overlap, there is a lower chance that the participant will hear speakers of the opposing dialect using low-activation variants that could cause such a large change, and instead a much higher chance that the participant will hear his partner use an exemplar that already falls within his general range of use, that is, one with a high activation. In that case, convergence towards this variant is very likely since the activation is already high, but for the very same reason, the variant is not so different that it causes a large magnitude of change if it is used. This is precisely the finding in the current study. It seems that speakers likely do stay within their phonetic repertoires when accommodating, but that the relevant repertoire is the entire perceptual set, rather than the set of production variants that makes up the speakers’ most common way of speaking. Furthermore, it may be the case that some of these previously only perceptual variants are used as production variants for the first time. To this extent, there seems to be something to the idea that speakers make greater changes on variables that have more “room” to be adjusted (Kim et al. 2011).
7.4 Gender

The model in §6.3.2.1 showed that there was a significant effect of gender on the likelihood of converging, with females more likely to converge than males. Females converged on 63.2% of the tokens, while males converged on 57.8% of tokens. Splitting this result out by dialectal difference, we see in Figure 51 that the female participants were more likely to converge on each of the differences than the male participants.

![Figure 51: Proportion of trials converged upon by gender and dialectal difference](image)

Previous studies have found conflicting results about whether men or women accommodate more. Pardo (2006) and Nielsen (2008) found that males were more likely to imitate than females, Bilous & Krauss (1988) and Namy et al. (2002) found that females accommodated more than males, and Babel (2009) and Black (2012) found that males and females imitated to the same degree. In Namy et al. (2002), 16 participants (8 women, 8 men) first read a list of 20 English words aloud to establish a baseline of pronunciation. Next, they shadowed those same words which had been pre-recorded by four different model speakers (2 women, 2 men). Lastly, 64 listener judges (32 women, 32 men) assessed imitation via an AXB task. Their results showed that female shadowers were more likely to imitate the model voices, and that female listeners were more likely to detect imitation. Namy and colleagues conclude that women’s greater likelihood to converge and to detect convergence reflect a greater perceptual sensitivity or greater attention paid to indexical information in the signal. If women are more sensitive to
such information then we might expect them to perform more accurately in the perception task in the current study. Indeed, as Figure 52 below shows, the female participants were more accurate at identifying the dialect of the tokens presented in the perception task than their male counterparts for each of the differences in the study; however, the difference in accuracy did not reach significance, possibly because there were only four female participants compared to 18 males.

![Figure 52: Proportion of trials correctly responded to in the perception task by gender and dialectal difference](image)

This result, although not statistically significant, seems to support the findings of Namy et al. (2002) that women may be more sensitive to indexical information in the signal, in this case, markers of dialect. If women are indeed more sensitive to this information, it might help to explain why they were more likely to converge.

### 7.5 Dialectal Difference #3: BAS Laminal [s] vs. MS Apical [ʂ]

Dialectal Difference #3 was expected to be fairly perceptually salient due to its inclusion in parodies (imitations) of Peninsular Spanish (Reiter 2004) and the fact that six of the participants mentioned it as a difference they had noticed between BAS and MS; however, the results of the perception task put the salience of this dialectal difference at 25% for BAS speakers and 22% for MS speakers, neither of which was higher than chance (20%). As explained in §6.2.3, the perceptual salience of Difference #3 may have been underestimated in the perception task due to
the fact that the stimuli used in the experiment only contained /s/ when the vowel /a/ followed, which may be one of the least salient contexts for this difference to be perceived, as compared to when /o/ or /u/ follow. In such contexts, the COG of /s/ is lowered, causing the MS /s/ to drop out of the range of COG produced by BAS speakers across all vowel contexts. As a result, Difference #3 may be more perceptually salient in those contexts. Since none of those contexts were tested, it seems that the perceptual salience of Difference #3 was not evaluated accurately, and it was for that reason that this difference was excluded from the rest of the data analysis.

If we examine the pattern of accommodation on Difference #3 here, both in terms of the magnitude of the change and the direction of the effect, we find that it patterns similarly to the highly salient dialectal differences. Figure 53 gives the mean absolute change made from pre- to post-conversation by dialect of the participants for Difference #3. The (standardized) absolute mean change is 1.09 for BAS speakers and 1.20 for MS speakers.

![Figure 53: The standardized magnitude of the change from pre- to post-conversation on Difference #3 by dialect](image)

In Figure 54, we see that the changes made on Difference #3 are higher than those made on all the other dialectal differences. A linear mixed effects model comparing the size of the change by dialectal difference and by dialect finds that the magnitude of the changes made on Difference #3 are statistically the same as those made on Differences #1 and #2 (which also do not differ from each other), all of which are higher than those made on the remaining differences (#4, #5,
and #6). In terms of the magnitude of the change then, Difference #3 patterns with the two most salient dialectal differences.

![Figure 54: The standardized magnitude of the change from pre- to post-conversation by dialectal difference by dialect](image)

Turning to the direction of the change, Figure 55 shows that the BAS speakers were more likely to diverge than their MS counterparts. In fact, the BAS speakers were more likely to diverge on Difference #3 than they were to converge (in terms of raw numbers; this difference is not statistically significant). This is the only dialectal difference which was more likely to be diverged on (for one of the dialects) than converged on. The BAS speakers converged on this difference only 49% of the time, while the MS speakers converged on it 66% of the time. A chi-square test found that this difference in likelihood of converging was significant ($\chi^2 = 11.99, p < 0.0001, df = 1$).
Why might the BAS speakers be more likely to diverge on /s/ than the MS speakers? In §7.2, I suggested that one of the factors that could account for asymmetries in the likelihood of converging on the two most salient dialectal differences (#1 and #2) was phonological contrast. The asymmetry found here for Difference #3 may also be explained by the need to maintain a phonological contrast, specifically, that such a need exists for BAS speakers, but not for MS speakers.

The difference in articulation between BAS and MS /s/ is reflected in the value of the centre of gravity of the /s/. Centre of gravity (COG) is an important acoustic cue for contrasts in place of articulation of coronal fricatives (e.g. Jongman et al. 2000), being negatively correlated with the size of the resonating cavity in front of the oral constriction. Thus, we expect, and found as discussed in §6.1.3, that BAS /s/ has a higher COG than the MS counterpart. BAS also has /ʃ/, another sibilant fricative, in its inventory and BAS speakers need to maintain this contrast, which is reflected in minimal pairs such as taza /ta.sa/ ‘cup’ and talla /ta.ʃa/ ‘size’. In the same way that the difference in anteriority between MS and BAS /s/ is characterized by a difference in COG, the phonemic contrast between BAS /s/ and BAS /ʃ/ is also characterized by a difference in COG. Since the place of articulation of /ʃ/ is more posterior than /s/, we expect that the COG of /ʃ/ will be lower than BAS /s/. Falling in between these two BAS phonemes is MS /s/, as illustrated in Figure 56 below.
Since the MS variant of /s/ falls in between a phonemic contrast in BAS, BAS speakers who converge towards the MS variant would be shrinking the size of the phonetic distance between the native /s/-/ʃ/ contrast, which could potentially threaten that contrast. This danger does not exist for the MS speakers, who have no other sibilant fricative in their inventories. If MS speakers converge upwards on COG towards the BAS variant of /s/, there is no contrast threatened. This asymmetry in the effect of converging on /s/ may explain why so many more BAS participants diverged on Difference #3 as compared to the MS participants.

To summarize, although we do not know the exact perceptual salience of Difference #3, its patterning in terms of the magnitude of the effect suggests that it is among the more perceptually salient differences included in the study. In addition, it seems that the asymmetry in likelihood of converging on /s/ between the dialects may be explained by appealing to a need to maintain a phonological contrast in the case of the BAS speakers, which is not present for the MS speakers.

### 7.6 Individual Differences in Accommodation

As discussed in §2, the pattern of accommodation can be affected by many different factors, including social, situational and linguistic factors. In a cross-dialectal interaction the perceptual salience of the differences and the need to maintain a phonological contrast also play a role.
Beyond these already numerous factors are characteristics of the individual speakers that can affect the pattern. Communication Accommodation Theory (CAT; Giles 1973) predicts that the direction of the change stems from a desire to minimize social distance or accentuate distinctiveness, which is, of course, a very personal and individual motivation. Studies of accommodation have found a number of other individual-level factors that could influence the direction and degree of change. As discussed in §2, Babel (2010) found that New Zealanders with a positive implicit bias towards Australia were more likely to imitate an Australian model talker. Similarly, Abrego-Collier et al. (2011) found that experiment participants were more likely to imitate a model talker if they evaluated the talker positively. Black (2012) found that participants who were high on emotional reactivity were more likely to imitate vowels than their counterparts who were low on emotional reactivity, but found the opposite trend for imitation of extended VOT. In addition, she found that the attractiveness of a model talker’s voice affected the imitation of VOT, but not in the expected direction, since speakers were less likely to imitate a voice that they preferred more than another.

Related to the effects of vocal attractiveness, Babel (2012) found that females who rated a White male model talker’s looks as more attractive were more likely to imitate the model talker’s vowels, but male participants who rate the model talker as attractive were less likely to imitate his vowels. Building on this finding, Babel et al. (2012) investigated the effects of attractiveness and prototypicality of model speakers’ voices. They found that both measures affected how often participants would imitate the model speakers’ voices. Both male and female shadowers were more likely to imitate atypical voices, in particular the least typical male voice. Females were more likely to imitate attractive voices (both male and female) than they were unattractive. Males showed the same pattern within the female voices, but for male voices they were actually more likely to imitate the least attractive voice than the most attractive male voice, supporting the findings of Babel (2012). These results suggest that not only do individual speakers have particular characteristics that make them more or less inclined to imitate another’s speech, people’s voices also have particular characteristics that may make them more or less likely to be imitated.

Although in the current study I did not explicitly measure how attractive the voices of the participants were to their partners, in the questionnaire that the participants completed after the experiment they were asked whether they liked the accent of the opposing dialect and which
of the two accents (their own or the opposing dialect’s) they preferred, another personal characteristic. In light of the findings discussed above, we might expect that participants who like the opposing dialect would be more likely to converge towards it, while those who do not like the opposing dialect would be more likely to diverge (or less likely to converge). There was variation in the participants’ responses to this question, with some saying they liked the other dialect and some saying they did not, although no participants responded that they preferred the opposing dialect over their own.

If we consider the participants who are the largest “convergers”, those who on average make the greatest positive changes towards the opposing dialect, we find that within the MS speakers there are 3 participants who converged much more than the other participants. These are the MS speakers in Pairs 1, 3, and 8, who made mean standardized changes towards the opposing dialect of 0.4511, 0.4707, and 0.4420, respectively. The mean amount of convergence for the MS speakers was 0.2636. All of these three participants indicated that they liked the BAS accent. The MS speaker in Pair 3 stated that while he still preferred the MS accent, the BAS accent was one of his favourites of all of the Spanish accents. The MS speakers were also asked whether they would want their Spanish to change if they one day lived in Argentina. In response, this participant said “En principio querría que no cambiara porque el acento es algo que caracteriza tu persona pero tampoco haría un drama si con el tiempo voy hablando con acento argentino”, meaning “In principle, I would not want it (my Spanish) to change because your accent is something that characterizes you, but it wouldn’t be a big deal if with time I came to speak with an Argentine accent”. The MS speaker in Pair 8 seemed more open to his Spanish changing saying “No me importaría, me parece natural, pero siempre hay cosas que quedan”, which means “It wouldn’t matter to me, it seems natural to me, but there are always things that remain”. So these MS participants who converged the most all stated that they liked the BAS accent and seemed at least somewhat open to their Spanish changing if they ever lived in Argentina.

The BAS participant who was the largest converger was the BAS speaker in Pair 10, making a mean standardized change of 0.4929, which was the largest of any of the participants. In his questionnaire he answered that he liked the MS accent, but preferred the BAS accent, which was a somewhat common response with 3 other BAS speakers saying the same; however, this participant also indicated that he believed his Spanish had changed since he had moved to
Madrid (3 months before) and that he used local words and verb formation. Furthermore, he noted that it did not matter to him if his Spanish changed and that he understand that it would likely change as a result of living in Madrid. The second largest BAS converger was the BAS speaker in Pair 1, making a mean standardized change of 0.2862, which is much smaller than that of the BAS speaker in Pair 10 discussed above, but still above the mean convergence by all of the BAS speakers (0.1985). This speaker was non-committal about whether or not she preferred the MS or BAS accent, saying that the accents of the regions were too diverse to be compared, but she stated that her Spanish had changed a lot since she had moved to Madrid and explained that she wanted her Spanish to evolve saying “Quiero que mi español esté vivo y para eso lo dejo fluir y en ese proceso habrá cambio que siempre es bueno porque habré aprendido un nuevo lenguaje”, meaning “I want my Spanish to be alive and for that I let it flow and in that process there will be change which is always good because I will have learned a new language”. These two speakers (from Pairs 1 and 10) showed the greatest amount of convergence towards the MS dialect within the BAS group and both of them expressed comfort with the idea of their Spanish changing.

This perspective, however, was somewhat rare among the BAS participants, who tended to say that they did not want their Spanish to change. One of the BAS participants who converged the least towards the opposing dialect (the BAS speaker in Pair 8, making a mean standardized change of only 0.0527, essentially maintenance) answered that he did not like the MS accent and preferred his own. He also explained that his friends and family had commented that his Spanish had changed although he did not notice it himself and did not want his Spanish to change saying “No quiero que cambie. Prefiero mantener la identidad de mi acento”, meaning “I don’t want it (my Spanish) to change. I prefer to maintain the identity of my accent.”

The MS speaker who converged the least (the MS speaker in Pair 7, making a change of -0.1046, a slight divergence on average) seemed to be open to the idea of his Spanish changing if he ever moved to Argentina saying, “Cambiaría seguro. Suelo adaptarme muy rápido a la musicalidad de las zonas donde viajo. Aunque seguramente volvería a recuperar el madrileño cuando volviera”, which means “It (my Spanish) would change for sure. I tend to adapt very quickly to the musicality of the areas where I travel. Although, I would certainly get the Madrileño (accent) back when I came back”. However, he also said that he did not like the BAS accent and was the only MS participant to say so. Seven of the remaining MS participants all
said they liked the BAS accent and 3 were non-committal. In my informal experience in Madrid, the BAS accent seemed very popular, with more than a handful of MS speakers (not only participants in the current experiment) remarking that they found it sexy, making the MS speaker in Pair 7’s opinion even more unusual. Furthermore, this MS speaker seemed somewhat mistrustful of Argentine people since when asked if he noticed any differences between the people of Madrid and Buenos Aires he answered “Sí, hay bastantes. El carácter es mucho más embaucador de los argentinos. En Madrid las personas somos más expresivas pero más directas también”, which means “Yes, there are lots (of differences). The character of the Argentine people is much more deceitful. In Madrid, the people are more expressive, but more direct as well”. So the speakers who converged the least commented that they did not like the accent of the opposing dialect and in some cases did not want their Spanish to change since they considered it a reflection of their identity.

While some of the answers given on the questionnaire might not be completely truthful due to participants potentially feeling the need to be diplomatic in their responses, we do see some interesting alignment of degree of convergence with attitudes towards the opposing dialect and desire to change or retain native dialect norms. These attitudes and desires are certainly an individual-level characteristic of the participants and it appears that they can have an important effect on the trajectory of phonetic accommodation. A similar finding was reported in Pesqueira (2008)’s study of the acquisition of the Mexican Spanish /j/ by Argentine speakers (to replace /ʃ/), which was discussed in §2.1.2.1, where various social factors, including the Argentine speakers having a positive attitude towards Mexican Spanish, were found to affect the proportional use of the Mexican glide.

Another factor found to be significant in Pesqueira’s study was whether or not the Argentine speakers planned to remain in Mexico or not. She found that those who planned to stay were more likely to take on the Mexican glide than those who did not plan to stay. The BAS participants in the current study were asked whether they planned to stay in Madrid in their questionnaire. Table 6 in §5.1 provided the results of that question. The BAS participants who were the largest convergers (the BAS speakers in pairs 1 and 10) both answered that they were planning to stay, while the smallest convergers (in pairs 4 and 8) answered that they were not planning to stay, providing further support for Pesqueira (2008)’s finding. Of course, the participants falling in between the largest and smallest convergers also either answered that they
were or were not planning to stay, so this factor is certainly not the only one at play, but our results do suggest that plans to remain in the D2 area may affect the extent to which speakers will converge towards the D2 dialect norms.

The majority of the factors discussed above relate to ‘liking’ of some characteristic of a model speaker, conversational partner, or new dialect; but, we might also find individual-level factors that are relevant to accommodation which have to do with differences in cognitive processing style. Recently, many studies have begun to look into the effects of differences in cognitive processing style as measured through the Autism-Spectrum Quotient (AQ) score. The Autism-Spectrum Quotient is a self-administered test developed by Simon Baron-Cohen and colleagues for determining the degree to which adults possess “autistic” traits. The test is made up of 50 statements, such as “I prefer to do things with others rather than on my own”, which participants respond to by indicating whether they “definitely agree”, “slightly agree”, “slightly disagree” or “definitely disagree”. The 50 statements cover 5 different areas: social skills, attention switching, attention to detail, communication, and imagination. A point is scored for each time a test-taker responds to a statement with the “autistic-like” behaviour mildly or strongly, meaning that higher scores indicate more autistic traits. Baron-Cohen et al. tested various groups of people and found that those with Asperger’s Syndrome or High-Functioning Autism had a mean AQ score of 35.8, which was significantly higher than the control group’s mean score of 16.4. Men were found to have a higher AQ score than women, and mathematics students were higher than science students, who were in turn higher than humanities students (Baron-Cohen et al. 2001).

The Autism-Spectrum Quotient has received some attention recently in research investigating the effect of variation in processing style (as measured by individuals’ AQ scores) on speech production and perception. People who would not be classified as having Autism or Aspergers Syndrome (corresponding to a high AQ score) can vary in their AQ score through having certain “autistic” traits, but not others. Building on the finding that individuals with Autism Spectrum Disorders may rely more on declarative memory than procedural memory (e.g. Ullman 2004), Mielke, Magloughlin & Nielsen (2012) investigated word-specificity effects in phonetic imitation for individuals with low and high AQ scores, with the prediction that individuals who rely on declarative memory more would be more likely to show the effects of word-specificity. In the study, participants read aloud a list of words before and after they were
exposed to recordings of a subset of those words for which initial-/p/ had been modified to have extended VOT. So the word list contained some /p/-initial words that the participants were not exposed to. The results found that speakers with high AQ showed a greater degree of imitation on the words they were exposed to than on those they were not exposed to, whereas the speakers with low AQ scores showed no difference.

Overall, it seems that differences in cognitive processing style, as measured via AQ score, may be related to differences in social characteristics, both of which appear to have effects on the pattern of accommodation in interactions. Although AQ score and measures of “liking” were not taken in the present study, it seems that these types of individual characteristics of the speakers may have significantly contributed to the extent of intra- and inter-speaker variability found in the pattern of phonetic accommodation.

7.7 The Mechanism Behind Accommodation

In the current study, salience was not defined as a purely linguistic concept (such as objective phonetic distance), but was defined within the context of how much particular linguistic variables contribute to the accurate recognition of a dialect. In §3, I argued that it is impossible to separate salience from the context in which linguistic variables are being perceived and still maintain salience as a meaningful concept. This means that the social meaning of dialects is always present while language users participate in interactions. This social meaning can inform part of our results, such as the fact that the most salient dialectal differences are likely consciously perceived by the participants who can use that knowledge to project their desired dialect identity. Nevertheless, it also seems to be true that linguistic factors, such as phonological contrast, play a role in affecting the degree and direction of accommodation. So in explaining the pattern of accommodation in the current study I appeal to both speaker-level social factors as well as linguistic, system-level constraints. To formalize the relative roles of these factors more clearly, we need to consider the mechanism behind accommodation.

In §2, I discussed the two main approaches in the literature to account for the mechanism behind accommodation. The social model, Communication Accommodation Theory (Giles 1973), posits that the driving force behind accommodation is speakers’ motivation to arrive at the desired social distance between themselves and their conversational partners. The automatic model assumes a direct connection between the perception and production systems and that
speech is perceived in terms of articulatory gestures that are, in turn, used as parameters for speech production, a relation that results in automatic and unintentional imitation (Pickering & Garrod 2004). As noted by Black (2012), these two models are not mutually exclusive since they seem to be related to two different aspects of the process of accommodation. The fact that imitation (or accommodation) has been found to occur even in non-social situations, such as rapid shadowing experiments in a lab setting (Goldinger 1998; Shockley et al. 2004; Mitterer & Ernestus 2008), seems to suggest the automaticity of the process. However, the fact that divergence is also attested along with variation in the degree of the change in either direction suggests that the process may be affected, if not initially driven, by social considerations, which may be best modeled using a framework like CAT, or by other linguistic or situational factors.

That both convergence and divergence on all of the variables and for both dialects were found in the current study seems to lend more support to the socially-conditioned hypothesis than to the automatic account, but like most patterns the reality probably falls somewhere in the middle. Convergence was more common than divergence, and the automatic account holds that convergence will be inevitable. As such, we might take the automatic account of the mechanism behind accommodation as the starting point to explain how accommodation comes about (via a direct link between perception and production), but then social, situational, and linguistic factors can mediate and affect the direction (and magnitude) of the change. This is the conclusion reached by Babel (2009) who writes “Neither a purely automatic nor a social theory that involves a conscious mechanism for accommodation is supported by the results of this experiment. Instead the data support an automatic theory of accommodation where we assume that implicit social factors will play a role on an unintentional level…” (Babel 2009: 129). It is not clear that the literature on CAT intends to propose that speakers must make conscious choices to converge or diverge. We could instead interpret these “choices” as low-level decisions not made consciously, but those that fall within the general sphere of behavioural convergence. The fact that accommodation (linguistic or non-linguistic) has social consequences suggests that speakers are aware of the process on some level, but it is not apparent that these choices must be conscious in order to draw a distinction between the automatic and social explanations for accommodation. Nevertheless, the spirit of Babel (2009)’s conclusion that imitation is an automatic process mediated by other factors (in her study, phonetic and social) does find support in the current study as well.
Convergence seemed to be the default behaviour with the majority of tokens being converged on taking all 22 participants as a whole. Each of the two dialects as groups also converged more often than they diverged. This pattern held for each dialect when split out by the six dialectal differences (except for the BAS pattern on Difference #3 /s/). Convergence was also more common than divergence at the individual level for 20 of the 22 participants. Even when splitting out the individual speaker pattern by the six dialectal differences we find that a full 71% of the speaker/difference combinations show convergence over divergence. This suggests that there is something more basic, or default, about convergence than divergence. Indeed, research has found that imitation is pervasive in many different behaviours including speech, facial gestures, and other body movements (Chartrand & Bargh 1999), and that imitation can be found in humans as young as newborn infants, as well as in other species (Meltzoff & Moore 1986; Ferrari et al. 2006). Of course, divergence is also frequent, and cannot be treated as a trivial departure from the basic pattern. I have appealed to both social factors, such as the desire to integrate within a new community or accentuate the distinctiveness of one’s own dialect norms, as well as linguistic factors, such as the need to maintain a phonological contrast, to explain the variation between convergence and divergence found as a result of conversation across two dialects of Spanish.

Cutting across this conclusion is the finding that perceptual salience has an effect on the pattern of accommodation as well. As discussed in §6.3.1.1, salience was found to affect the magnitude of the changes made from pre- to post-conversation in that the participants made greater changes as perceptual salience increased. Furthermore, this effect was consistent across all 22 participants, suggesting that the reaction to salience (with respect to the magnitude of the change) is something automatic, whereas the direction of the change (convergence or divergence) did not have a consistent relationship with the salience of the differences. As we saw in §6.3.2.1, the by-speaker random coefficients for %-CORRECT varied between positive and negative slopes and in terms of the steepness. If the direction of the change is predicted by social factors, rather than a more system-level mechanism, then it is logical that these slopes showed no pattern across the participants.

This finding that perceptual salience affects the magnitude of the change, but that the direction of the change is determined primarily by social factors and phonological contrast restrictions is one of the main conclusions and contributions of this study.
7.8 The Asymmetry of Convergence and Divergence

By definition, convergence and divergence are mirror images of each other: when a speaker converges to another he changes his speech to become more similar to his interlocutor, and when he diverges he changes his speech to become less similar to his interlocutor. While the likelihood of convergence or divergence to occur seem to be equal (the myriad social and linguistic factors that can affect the likelihood of one or the other notwithstanding), the potential magnitude of a convergence and divergence may not be equal. If two speakers differ in their production of a shared segment by a certain amount, one of the speakers could potentially change his realization by 100% of that difference in order to perfectly align his production with his interlocutor’s. The maximum potential for convergence (ignoring the possibility of over-convergence) is equal to 100% of the distance between the two speakers before they engage in conversation. Conversely, if a speaker were to diverge by the same amount (diverging by 100% of the difference) he could very likely end up producing a completely different segment. For example, if a BAS and MS speaker differed in their pre-conversation production of /s/ by 1000Hz (with the MS speaker being the lower of the two), the MS speaker could raise his centre of gravity by that 1000Hz to fall exactly into line with his BAS counterpart, thereby converging completely. If he were to diverge by 1000Hz, however, lowering his COG by that amount, he could end up producing something more akin to the postalveolar fricative [ʃ]. Although this would technically be a divergence on the acoustic measure that distinguishes BAS and MS /s/, such a strategy would likely introduce difficulties in comprehension on the part of the BAS speaker, who would almost certainly perceive the postalveolar fricative as a realization of orthographic <ll> or <y> and who would wonder why an MS speaker was using that sound at all. In this way, it seems that there is an asymmetry in the potential for convergence and divergence with respect to the magnitude of the change. The potential magnitude of a convergence is greater than the potential magnitude of a divergence.

In the current study it was found that the mean magnitude of the change made on tokens where convergence occurred was statistically significantly greater than on the tokens where divergence occurred (see §6.3.1.1). This difference may stem from the asymmetry in the potential magnitude of convergence and divergence. One unit of convergence may not be equal to one unit of divergence. One possible way to deal with this is to normalize the magnitude of the changes across convergence and divergence by incorporating a perceptual measure. If native
speaker judges perceive a particular magnitude of divergence more often (or as greater) than the same magnitude of convergence, then we could potentially use this pattern to normalize the size of convergences and divergences perceptually. Many studies of accommodation have used the perceptions of native speaker judges to assess the degree and direction of change, as discussed in §2.

7.9 Salience and Non-Phonetic Accommodation

One of the main findings of the current study is that the magnitude of phonetic accommodation increases as the perceptual salience of the dialectal differences increases. Given this finding, we might expect that perceptual salience would play a similar role in accommodation when considered across different domains of speech, beyond only pronunciation. The most salient domain is likely the lexicon, where the use of dialect-specific lexical items would be highly salient. We might expect the participants who converged to a greater extent and more often on the more salient phonetic differences to show a similar effect of perceptual salience across domains by taking on the use of D2 lexical items more than other participants who are less inclined to converge in the phonetic domain.

The BAS participant who converged the most was the BAS speaker in pair 10. This participant was discussed in §7.6 in relation to his attitudes towards the MS dialect and with respect to his openness to having his Spanish change as a result of living in Madrid, as reported on his questionnaire. As mentioned in that section, this participant said that it did not matter to him if his Spanish changed and that he understood that it would likely change as a result of living in Madrid. He also stated that he was already using local lexical items and verb formation, suggesting that he had taken on the use of the second person singular informal pronoun tú and the third person plural pronoun vosotros (along with their corresponding verb morphology), instead of the vos and ustedes forms that he would have used in Argentina. This suggests that perhaps this participant is sensitive to salience beyond just the phonetic domain, accommodating on lexical items and verbal forms, which are likely highly salient. No other participants specifically mentioned taking on lexical items or verb forms in their questionnaire so it is difficult to compare this high converger to a low converger (or diverger) in terms of how the participants reported their use of local lexical items.
However, to verify that the BAS speaker in pair 10 actually did show some evidence of having accommodated to MS lexical items and the use of the Peninsular Spanish pronouns and verb forms, I listened to the conversation he had with his MS partner, focusing on the use of these pronouns, as well as other potential MS-specific lexical items, such as the ubiquitous *vale*, meaning ‘ok’. In the map task where this participant was giving instructions to his partner (providing many opportunities for him to exemplify his use of *tú* or *vos*, as well as the second person singular imperative form which also differs between the two dialects), the BAS speaker produced 68 verbs in the second personal singular informal conjugation. Of those 68, 32 were realized using the BAS verb form (47%) and 36 were realized with the MS form (53%), with several verbs being produced at times using the BAS form and at others with the MS form. For example, at one point the participant provides the instruction *Te dirigis hacia la izquierda* ‘you head to the left’, using the BAS form of the verb *dirigirse* ‘to go/head’. After his partner responds, the BAS speaker tells him *Luego, te diriges hacia abajo* ‘then you head downwards’, using the MS realization of the very same verb. So this participant, who was the largest converger of all of the participants, seems to use the MS and BAS verb forms about equally. Although he mentioned that he also used local lexical items, there was not a single instance of the word *vale*, which was included very frequently in the MS participants’ speech.

If the salience of cross-dialectal differences in the lexicon is expected to affect participants the same way that salience in cross-dialectal phonetic or phonological differences do, we might expect that if the BAS participant who converged most also uses the MS verb forms about half the time, perhaps the BAS participants who converged the least would use the MS verb forms less. The two BAS participants who converged least were the speakers in pairs 4 and 8. The speaker in pair 4 used BAS forms in 39% of possible instances, and the MS forms in 61%, showing that the MS forms were used more than for the largest BAS converger. A more extreme result was found for the second smallest BAS converger: the BAS speaker in pair 8. He used the BAS verb forms only 3 times (10%) while instructing his partner, while using the MS forms the remaining 90% of the time.

So the largest BAS converger used the MS verb forms *less* often than the smallest BAS convergers. While this seems to show the opposite of what was predicted above, in the cases of both the BAS speakers in pair 4 and 8, there are circumstances that make drawing conclusions about the use of MS verb forms difficult. The BAS speaker in pair 4 used a very small range of
verbs when instructing his partner, focusing almost exclusively on tener que ‘to have to’, and pasar ‘to pass’. For example, he frequently gave instructions like Tienes que bajar hacia el restaurante ‘You have to go down towards the restaurant’ or Luego pasas por la derecha de la escuela ‘Then you pass by the school on the right’. In contrast, the larger converger, the BAS speaker in pair 10, used many different verbs during his conversation. While the BAS speaker in pair 4 used the MS forms more often, it is not possible to determine whether this use would be found in many verbs, or whether it is concentrated in the few he used in the recording. In the case of the BAS speaker in pair 8 (the second smallest BAS converger), a similar result was found in that he too used tener que very frequently and always with the MS morphology. In addition, the majority of the verb forms he produced were in the second person plural (nosotros), where there is no difference in realization of the verbs between the dialects.

In all of these cases, the use of the MS verb forms might not be evidence that the BAS speakers have actually acquired these forms, but rather may reflect the fact that, as instruction givers in the map task, they were attempting to make the instructions as clear as possible for the MS speakers. The more free-form conversation component, in which the participants spoke about the differences between Madrid and Buenos Aires and things they liked about the cities, provided far fewer opportunities to hear differences in second person verb forms. Of course, there are other lexical differences between the dialects besides the verbs; however, the conversation component was not designed to elicit specific words, instead focusing on providing exposure to the pronunciation differences between the dialects.

Overall, taking these results at face value, we see little evidence to suggest that salience across linguistic domains will affect participants the same way that salience within the phonetic characteristics of sounds does. This could be due to differences in how linguistic information is stored and indexed socially in the various domains, potentially affecting how salience is able to interact with accommodation. In addition, it may be that differences in when lexical replacements and pronunciation changes are acquired during acquisition of a second dialect preclude a meaningful investigation of both processes at only one time period. In Chambers (1992)’s discussion of his 8 principles of dialect acquisition, his first principle is that lexical replacements are acquired faster than pronunciation and phonological variants. Furthermore, his second principle states that lexical replacements will take place rapidly at the first stage of dialect acquisition and then slow down. This is at least partly due to potential difficulties in
comprehension stemming from lexical items that are not found in the D2, or that have a different meaning in each dialect. If lexical replacements are expected to take place before changes in pronunciation, then perhaps it is not unexpected that even the BAS speakers who converged very little with respect to pronunciation both showed high usage of MS verb forms (at least in the verbs they produced). For these speakers, it is still possible for changes in pronunciation to take place further into the process of dialect acquisition.

In sum, it seems that the finding of this study that as perceptual salience increases so does the magnitude of the change does not necessarily extend to consideration across domains of speech.

7.10 Salience, Accommodation, and Sound Change

Accommodation within a single interaction is the first stage of the Change-by-Accommodation model (Niedzielski & Giles 1996), discussed in §2, with two stages following the first: 2. long-term dialect accommodation resulting in permanent changes in the speech of individual speakers, and 3. the spread of new speech habits throughout the community. Pertaining to the first stage, the most important finding of this thesis is that within a conversation between speakers of two different dialects of Spanish, the pattern of phonetic accommodation is affected by the perceptual salience of the dialectal differences in that the most salient differences are the most subject to change. In this section, I explore the predictions that this finding makes for the second and third stages of the Change-by-Accommodation model, as well as for patterns of sound change more generally.

Given that the participants in the current study made the greatest changes on the most salient dialectal differences, we would expect that over time, the changes that they make would accumulate faster on these particular dialectal differences, resulting in faster or more complete acquisition of these variables. Of course, we have seen that phonological contrast can affect the direction of change, so the prediction should be altered to say it would be the most salient variables that did not threaten a D1 phonological contrast that would be acquired first. This prediction is focused on those speakers whose social motivation is to integrate with the new community and minimize social distance between themselves and the D2 speakers. On the other hand, the speakers who do not wish to integrate or who wish to accentuate their native dialect norms would also be expected to make the largest changes on the most salient variables, but we
would not expect them to converge on these variables, rather, they may diverge, as they did in the short term, or we might see that these speakers would simply maintain their original pronunciation of these variables over time.

The Change-by-Accommodation model predicts that the acquisition of D2 features at the individual level is the precursor to the spread of these habits throughout the community, resulting in a larger scale change. Given the results of this study, we might expect that the longest term changes will be skewed towards convergence since, as found in §6.3.1.1, the magnitude of the change was significantly larger when speakers converged than when they diverged (mean standardized absolute change for convergence: 0.885, and for divergence: 0.695). In addition, it was found that convergence is more common than divergence (59% of tokens were converged upon, 41% diverged upon). Over time then, the largest changes that speakers will be exposed to will be convergences and speakers will be exposed more often to convergence. In the cases where a phonological contrast was threatened, we might expect that the progress of community level change (likely a levelling of differences between the dialects) will proceed more slowly than the instances where it is not. In this way, we can see that the findings of the initial stage of the Change-by-Accommodation model predict a specific outcome at both the second and third stages as well. Of course, these outcomes assume no major changes in how perceptually salient the dialectal differences are for the speakers involved in the contact situation as well as no changes in the speakers’ disposition towards the opposing dialect.

The predictions of the Change-by-Accommodation model are focused specifically on sound changes that take place within situations of dialect contact. Shifting to diachronic sound change more generally, Yu (2010) recently made a connection between AQ score, discussed in §7.6, and the robust finding that women are typically the leaders in sound change (e.g. Labov 1990, 2001). In that study, Yu investigated variation in perceptual compensation stemming from differences in cognitive processing style measured via AQ score. He specifically investigated perceptual compensation for vocalic context and talker voice in an English /s/-/ʃ/ continuum and how compensation differed as a function of AQ score. He found that females with a low AQ score compensated less when the vowel /u/ followed the sibilant (showing a greater likelihood of perceiving /ʃ/ due to a lowering of the second formant resulting from rounding of the following vowel) than those with a high AQ score. This effect was not found for men. This finding indicates that there is a greater potential for misperception of /s/ as /ʃ/ for low AQ
females than for high AQ females or for males. However, Yu found no difference in perceptual compensation between the genders for talker voice, meaning that there is no difference in the potential for misperception based on talker voice between the genders or depending on AQ score. These results are interesting because, as Yu notes, sound changes that stem from differences in talker voice are unattested while sound changes stemming from misperceptions due to vocalic context abound (e.g. Ohala 1993). And, it is within the perceptual compensation for vocalic context where Yu finds that females with low AQ score show an increased likelihood of misperception. Females have been shown many times to be the leaders in innovating and propagating sound changes (e.g. Labov 1990, 2001). Yu’s result suggests that part of the reason that females come to have innovative variants in their production may be cognitive, since it may be their low AQ score that causes the lack of perceptual compensation, which in turn introduces new variants into unintended phonetic categories. The finding may be linked to Ohala’s theory of sound change, in which “mini sound changes” result from “innocent” misperceptions (Ohala 1987, 1993). Perhaps part of the reason why such misperceptions occur stems from differences in cognitive processing style.

Speakers who propagate a sound change are known as early adopters, while those who generate the change at its earliest stage are the innovators. It is well established that women tend to lead sound change by being the ones to propagate new variants throughout the community, but does this mean that these same women are the ones who are the innovators of the new variants? As noted by Denis (2011), finding the innovators of a linguistic change is frequently impossible since we cannot know about a change before it has already begun. In his study, Denis makes use of the York English Corpus (Tagliamonte 1999-2001), which captures the onset of a linguistic change involving the general extender *stuff*, to investigate the social characteristics of the speakers who are the innovators of this development. He concludes that they may indeed be the same people who are the early adopters of the variable since many of their social characteristics overlapped with those provided by Labov (2001) as pertaining to early adopters. So, if low AQ females are the ones who create a new variable (because of lack of perceptual compensation) then we may also expect that they will be the same ones to begin propagating that change, at least in certain situations of sound change based on phonetic context. Chambers (2003) says that females are the ones to propagate a linguistic change because they have dense networks and are upwardly mobile, so they are in contact with more people and they
have more chances to be exposed to different variants and have more chance to expose others to the variants they have imitated, as well as those they have generated.

How does this picture relate to accommodation? If low AQ females are more likely to generate and propagate a change, then we might expect that they would be more likely to imitate other speakers and to pass on their new variants to others in the community. As discussed in §2, previous studies have found conflicting results about whether men or women accommodate more, with some finding that women accommodate more (Bilous & Krauss 1988; Namy et al. 2002), others finding that males accommodate more (Pardo 2006; Nielsen 2008), and still others finding no difference in accommodation between the genders (Babel 2009; Black 2012). As we saw in §6.3.2, in the current study the female participants were more likely to converge than their male counterparts, a pattern which held across all of the dialectal differences.

There was no measure of AQ score taken in this experiment, but females do tend to have lower AQ score than males (Baron-Cohen et al. 2001). In addition, studies have suggested that AQ scores are associated with different personality traits, with higher AQ scores being related to lower agreeableness and extraversion, and higher neuroticism (e.g. Austin 2005). It seems reasonable that individuals who are evaluated as less agreeable and less outgoing with higher neuroticism might have more difficulty managing interpersonal relationships and interactions. Connecting this to accommodation, Street & Murphy (1987) found that males with low interpersonal orientations were less likely to converge towards their conversational partner than males with high interpersonal orientations. Additionally, Pardo et al. (2012) found a significant correlation between convergence and reported closeness between roommates, where roommates who had a closer relationship converged towards each other more. If stronger relationships result in greater convergence, then we would expect that the speakers who are more able to generate strong relationships will be those are most likely to show convergence either within a single conversation or over time.

In a community where only one dialect of a particular language is spoken, we may expect that (low AQ) females will be the ones to introduce new variants based on lack of perceptual compensation, as well as to be the early adopters of new variants introduced by others, and then to be the ones to propagate this change throughout the community via their dense networks and perhaps, particular social characteristics that make them more likely to
imitate and be imitated by others. In such a situation, the dialect of that community changes over time, with no particular pre-defined direction. In a community where two dialects are spoken side by side, we can expect a similar situation, but with an added layer of complexity where new variants may be added to the speech of speakers via differences between the two dialects. That is, females may be expected to add new variants not only that stem from a lack of perceptual compensation (or other difference in processing), but also through exposure to D2 variants. In this situation, as D2 variants are propagated (by female speakers of both dialects), there is a direction to the change, where the differences between the two dialects are levelled, arriving at some intermediate point between the two. Once a stable point is reached, sound change can continue following the undirected path expected with a single dialect community.
Chapter 8
Conclusions

The main goal of this thesis was to investigate the effect of perceptual salience on the pattern of cross-dialectal phonetic accommodation in conversation. To achieve this goal, a metric of the salience of the differences between the dialects was developed and employed, based on the arguments made in §3 that salience is a measurable property of linguistic variables. Furthermore, in §3, I argued that salience must be measured and presented evidence of the flaws of the criteria-list approach to salience. The perception task designed to measure salience found that within the six dialectal differences I included, there were two levels of salience for the BAS speakers (high and low) and three levels of salience (high, moderate, and low) for the MS speakers. With the measurements of salience in hand, I then examined the degree and direction of changes that BAS and MS speakers made on these six dialectal differences as a result of engaging in a conversation with a speaker of the other dialect.

The main finding is that perceptual salience affects the magnitude of the change since the greater the salience, the larger the change. There was no consistent effect of perceptual salience on the direction of the change. To some extent, these findings support Trudgill (1986: 11) who predicted that speakers of a D1 in contact with speakers of a D2 would show the greatest amount of convergence on the most salient dialectal differences. The most salient differences in the current study did show the largest changes, but these changes were not always convergence. I argued that the direction of the change is determined primarily by the social motivations of the individual speakers. Those who wish to minimize social distance between themselves and their interlocutors, or who wish to integrate with the D2 may be more likely to converge, while those who wish to accentuate distinctiveness and maintain their D1 norms may be more likely to diverge. The perceptual salience of the dialectal differences was suggested to affect the magnitude of the change because, whichever motivation a speaker has, making a large change on the most salient differences would likely make the greatest impact in terms of their perceived convergence or divergence.

Overlying these results was the effect of the need to maintain a phonological contrast. This effect was relevant for the BAS speakers in terms of Difference #3 /s/ in that for them, to converge towards the MS /s/ could potentially threaten the BAS /s/-/ʃ/ contrast. As a result,
divergence was more common on this difference than convergence for the BAS speakers. Likewise, phonological contrast appeared to play a role in explaining an asymmetry in the direction of change between the dialect for the two most perceptually salient dialectal differences (Differences #1 and #2), which differed in whether or not they represented a contrast in each of the two dialects.

Present at both the inter- and intra-speaker level was much variation, as expected. I appealed to various individual-level factors, most notably the social motivations best captured under Communication Accommodation Theory (Giles 1973) to minimize social distance or accentuate distinctiveness, factors related to “liking” of some aspect of the model speaker such as his voice or his dialect, as well as potentially cognitive factors such as those reflected in the Autism-Quotient score to try and explain some of the variation.

Although certainly not the main goal of this thesis, the results also weighed in on the mechanism behind accommodation and suggested that a direct link between perception and production couched with an Exemplar Theory framework (Johnson 1997; Pierrehumbert 2001) may explain the initial trigger for convergence, but that the process was mediated by various social and linguistic factors, including the perceptual salience of the dialectal differences. Furthermore, I found that an exemplar approach in which the participants have all of the dialectal variants in their perceptual pool, and thus have access to them for production targets as well, may be able to account for the relationship between overlap and the magnitude of the change and between overlap and the direction of the change.

While the body of research examining accommodation (phonetic or otherwise) is growing, there remain many questions to address. Within the existing literature this thesis is unique in that it has taken on several aspects of the pattern of phonetic accommodation that have not been considered explicitly before. First, the examination of perceptual salience and its role in accommodation constitutes the main contribution. In this study, I have presented evidence that perceptual salience does play a role in accommodation, and that merely stating how salient a linguistic variable is may not accurately reflect the perceived salience of that variable for individual speakers. Furthermore, I have shown that how perceptually salient a particular variable will be is intricately linked with the context in which the variable is perceived, that is, the social meaning that is attributed to the use of that variable. In this study, I have examined the
salience of the dialectal differences within the context of dialect recognition. In a study interested in the effect of social class or ethnicity, the measurements I have presented for these dialectal differences would likely prove irrelevant. The point is that the context in which linguistic variables are perceived affects the social meaning that listeners receive and the social characteristics that they will attribute to a speaker who uses a particular variable. What is salient for dialect recognition might not be salient for identification of ethnicity (as found in Torbert 2004). Moreover, by using an experimental technique to quantify salience I have shown that while there are certainly patterns in which variables are considered more or less salient to the participants, there was also much inter-speaker variation in accuracy, indicating that we must not assume that all speakers of a given dialect will perceive dialectal variants in the same way, as assumed by the criteria-list approach to salience.

The second contribution of this thesis is that it focused on accommodation in Spanish, whereas the vast majority of studies on short-term accommodation have worked with English (although see Kim et al. 2011’s study on accommodation in Korean). It is not surprising that accommodation occurs in Spanish as well as in English, especially if the mechanism behind accommodation is a direct link between the perceptual and production systems, which would be common to speakers of all languages. However, it is important to study accommodation in various languages and cultures, since the social motivations to change may differ dramatically depending on the social, economic, and political climate in which the speakers find themselves.

A third contribution of the thesis is that it considered accommodation occurring across two different dialects. Research on cross-dialectal accommodation is growing and represents an important area since the factors that may affect accommodation in this situation are more nuanced and complex than in a single dialect situation. In particular, issues of the projection of identity as a native of a particular region may be paramount, depending on the specific inclination of the individual. Such factors play less of a role in single dialect conversation. Understanding the factors at work in cross-dialectal conversation helps us make predictions regarding the later stages of dialect contact, including the acquisition of a second dialect by individual speakers as well as the trajectory of changes that we might expect long term within dialect levelling or mixing.
With any study there are problems or limitations associated with the particular methodology employed. In the current study the main limitation stems from the fact that changes made on six dialectal differences cued by various acoustic measures needed to be compared. How do you compare a change in centre of gravity with a change in the duration of a vocalic sequence? How can we know which of these changes is greater? Normalizing across these changes proved one of the greatest challenges (read: headaches) of this study. The most intuitive method is to calculate the size of the difference between two speakers (or between a speaker and the opposing dialect) on a particular measure in the pre-conversation production session, calculate the change the participant makes from pre- to post-conversation, and then generate a percentage of the pre-conversation gap, reflecting the percentage that he has closed (or widened) that gap. To say that a participant converged 50% towards the opposing dialect indicates that the gap between his production of a particular variable and the other dialect’s production is 50% smaller than it was before he had the conversation with a speaker of the other dialect. This seems very intuitive; however, the problem with this method is that for dialectal differences that are very small to begin with, a participant need only make a very small (perhaps meaningless) raw change in order to generate a huge percentage change. With this method, the smaller the difference in pre-conversation, the greater the potential for large changes by post-conversation. The potential for the magnitude of the change needs to be equal across all of the dialectal differences. It is for this reason that I calculated the raw change from pre- to post-conversation for each token and then standardized the changes relative to each speaker’s pre-conversation production of each dialectal difference so that the changes could be directly comparable across the differences and across speakers: a 0.8 change on Difference #1 (measured via relative intensity) is thus equal to a 0.8 change on Difference #6 (measured via duration of the vocalic sequence).

Of course, these changes are “equal” in the sense that the units with which I am comparing them are now the same. They have been standardized from different acoustic units that reflect some reality in the production. It is not clear that the perceptual effect of a 0.8 change on Difference #1 would be equivalent to the perceptual effect of a 0.8 change on Difference #6. Listeners likely do not weight changes made on different acoustic measures equally, as suggested by Evans & Alshangiti (2010). This is one of the reasons that many researchers have chosen not to perform acoustic analysis in their studies of accommodation or imitation, but instead to conduct an AXB perception task where native speaker judges decide...
whether a pre-exposure production is more similar to a model talker than a post-exposure production. The downside to the AXB method is that it is not clear which acoustic information in the signal the listeners are using to make their judgements. In the current study, it was crucial to be able to explore the changes in particular acoustic measures since the dialectal differences are reflected in such measures.

Building on the results found in this thesis there are several avenues for future research. First, with the pattern in short-term accommodation identified, it would be interesting to track changes within several individuals over time, both in terms of how their changes on particular dialectal differences continue, as well as how their perception of the salience of those differences might change as well. In this way, we could determine to what extent the short-term pattern is predictive of longer-term patterns. A somewhat different angle concerns how native speakers of a dialect might perceive the speech of individuals who have lived for varying amounts of time in another dialect area. As noted by Munro et al. (1999), while anecdotal evidence of friends or family members living abroad and coming back with a “funny” accent is common, such evaluations may not be based on actual changes in production towards the norms of the D2 or they may be made post-hoc, in that a listener knows that his friend has lived abroad and thus, expects him to sound different. Relating to the perceptual salience of dialectal variants, we may find that which particular variants a transplanted speaker takes on may affect the extent to which naïve listeners judge his pronunciation on a scale of D1 to D2.

Our understanding of the complexities of phonetic accommodation in conversation is developing as more and more researchers investigate the various social, situational, and linguistic factors that affect the process. This thesis contributes to this development by exploring how the perceptual salience of dialectal differences interacts with the changes that speakers make in conversation with a speaker of a different dialect.
References


Babel, Molly (2009). *Phonetic and social selectivity in speech accommodation*, PhD dissertation, University of California, Berkeley


Black, Alexis (2012). Acoustic and social parameters on phonetic imitation: gender, emotion, and feature saliency, qualifying paper, University of British Columbia


Brown, LeAnn (2011). The integration of acoustic cues in the perception of gender and sexual orientation, qualifying paper, University of Toronto

Brunner, Elizabeth (2010). Imitation, awareness, and folk linguistic artifacts, PhD dissertation Rice University


Carreira, Maria (1998). A constraint-based approach to Spanish spirantization. Theoretical Analysis on Romance Languages: selected papers from the 26th Linguistic Symposium on Romance Languages. Mexico City: John Benjamins. 143-157

Cebrian, Juli (2002). Phonetic similarity, syllabification and phonotactic constraints in the acquisition of a second language contrast, PhD dissertation, University of Toronto


Denis, Derek (2011). Innovators and innovation: tracking the innovators of and stuff in York English. *University of Pennsylvania Working Papers in Linguistics* 17(2)


Face, Timothy & Scott Alvord (2004). Lexical and acoustic factors in the perception of the Spanish diphthong vs. hiatus contrast. *Hispania* 87(3): 553-564


Heffernan, Kevin (2007). *Phonetic distinctiveness as a sociolinguistic variable*, PhD dissertation, University of Toronto


Hwu, Fenfang (1994). *Consonant weakening and strengthening in Spanish dialects*, PhD dissertation, University of Illinois at Urbana-Champaign


Labov, William (1990). The intersection of sex and social class in the course of linguistic change. Language Variation and Change 2: 205-254


Ladefoged, Peter & Ian Maddieson (1996). The sounds of the world's languages, Cambridge, MA: Blackwell

Laferriere, Martha (1979). Ethnicity in phonological variation and change. Language 55(3): 603-617


MacLeod, Bethany (2007). Variable realization of vocalic sequences across Spanish dialects. MA thesis, University of Toronto
MacLeod, Bethany (2010). Variable velar lenition and the perception and production of [w] in Buenos Aires Spanish. Paper presented at Experimental Approaches to Perception and Production of Language Variation (ExApp 2010). University of Groningen, Groningen, The Netherlands


Makashay, Matthew Joel (2003). Individual differences in speech and non-speech perception of frequency and duration, PhD dissertation, The Ohio State University


Mazzaro, Natalia (2011). Experimental approaches to sound variation: a sociophonetic study of labial and velar fricatives and approximants in Argentine Spanish, PhD dissertation, University of Toronto


Nielsen, Kuniko (2008). Word-level and feature-level effects in phonetic imitation, PhD dissertation, University of California, Los Angeles

Ohala, John (1987). Sound change is drawn from a pool of synchronic variation. In Proceedings of Symposium on "The causes of language change, do we know them yet?", University of Tromso, Norway. 15-17


Quilis, Antonio (1981). Fonética acústica de la lengua española, Madrid: Gredos


Roca, Iggy (1997). There are no "glides"; at least in Spanish: an Optimality account. Probus 9: 233-265


Sadlier-Brown, Emily (2011). Children, Canadian Raising and new directions for an old variable. Paper presented at Change and Variation in Canada V. University of Victoria


Tagliamonte, Sali A. (1999-2001). Grammatical variation and change in British English: Perspectives from York. Economic and Social Science Research Council (ESRC) of Great Britain


Appendices

Appendix A – Stimuli

Dialectal Difference #1 (BAS /s/ vs. MS /θ/): orthographic <z>

Table 23: Stimuli for Dialectal Difference #1: orthographic <z>

<table>
<thead>
<tr>
<th>vowel</th>
<th>position</th>
<th>stress</th>
<th>token</th>
<th>BAS</th>
<th>MS</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>initial</td>
<td>yes</td>
<td>zaga</td>
<td>[sa.ya]</td>
<td>[θa.ya]</td>
<td>‘defense’</td>
</tr>
<tr>
<td></td>
<td>initial</td>
<td>no</td>
<td>zapatería</td>
<td>[sa.pa.te.'ri.a]</td>
<td>[θa.pa.te.'ri.a]</td>
<td>‘shoestore’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>realizar</td>
<td>[re.a.li.'sar]</td>
<td>[re.a.li.'θar]</td>
<td>‘to carry out’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>no</td>
<td>plaza</td>
<td>[pla.sa]</td>
<td>[pla.'θa]</td>
<td>‘plaza’</td>
</tr>
<tr>
<td>/e/</td>
<td>initial</td>
<td>yes</td>
<td>zeta</td>
<td>[se.ta]</td>
<td>[θe.ta]</td>
<td>‘z’ (letter)</td>
</tr>
<tr>
<td></td>
<td>initial</td>
<td>no</td>
<td>zelanda</td>
<td>[se.lan.da]</td>
<td>[θe.lan.da]</td>
<td>‘(New) Zealand’</td>
</tr>
<tr>
<td>/i/</td>
<td>initial</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>initial</td>
<td>no</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/o/</td>
<td>initial</td>
<td>yes</td>
<td>zona</td>
<td>[so.na]</td>
<td>[θo.na]</td>
<td>‘zone’</td>
</tr>
<tr>
<td></td>
<td>initial</td>
<td>no</td>
<td>zoológico</td>
<td>[so.o.'lo.xi.ko]</td>
<td>[θo.o.'lo.xi.ko]</td>
<td>‘zoological’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>razón</td>
<td>[ra.'son]</td>
<td>[ra.'θon]</td>
<td>‘reason’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>no</td>
<td>pozo</td>
<td>[po.so]</td>
<td>[po.'θo]</td>
<td>‘well’</td>
</tr>
<tr>
<td>/u/</td>
<td>initial</td>
<td>yes</td>
<td>zumo</td>
<td>[su.mo]</td>
<td>[θu.mo]</td>
<td>‘juice’</td>
</tr>
<tr>
<td></td>
<td>initial</td>
<td>no</td>
<td>zumbido</td>
<td>[sum.'bi.ðo]</td>
<td>[θum.'bi.ðo]</td>
<td>‘buzz’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>azul</td>
<td>[a.'sul]</td>
<td>[a.'θul]</td>
<td>‘blue’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>no</td>
<td>dulzura</td>
<td>[dul.'su.ða]</td>
<td>[dul.'θu.ða]</td>
<td>‘sweetness’</td>
</tr>
</tbody>
</table>
### Table 24: Stimuli for Dialectal Difference #1: orthographic <c>

<table>
<thead>
<tr>
<th>vowel position</th>
<th>stress</th>
<th>token</th>
<th>BAS</th>
<th>MS</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial</td>
<td>yes</td>
<td>cerca</td>
<td>['səɾ.kə]</td>
<td>['θəɾ.kə]</td>
<td>‘close’</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>centro</td>
<td>['sen.tɾo]</td>
<td>['θen.tɾo]</td>
<td>‘centre’</td>
</tr>
<tr>
<td>/e/</td>
<td></td>
<td>cerveza</td>
<td>['sɛɾ.ˈβe.sa]</td>
<td>['θɛɾ.ˈβe.θa]</td>
<td>‘beer’</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>cervecería</td>
<td>['sɛɾ.βe.ɾe.ɾi.a]</td>
<td>['θɛɾ.βe.ɾe.ɾi.a]</td>
<td>‘bar’</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>hacer</td>
<td>[a.ˈser]</td>
<td>[a.ˈθer]</td>
<td>‘to do’</td>
</tr>
<tr>
<td>medial</td>
<td>yes</td>
<td>proceso</td>
<td>[pro.ˈse.so]</td>
<td>[pro.ˈθe.so]</td>
<td>‘process’</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>entonces</td>
<td>[en.ˈton.ɾes]</td>
<td>[en.ˈθon.ɾes]</td>
<td>‘then’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>veces</td>
<td>[ˈbe.ɾes]</td>
<td>[ˈbe.θes]</td>
<td>‘times’</td>
</tr>
<tr>
<td>initial</td>
<td>yes</td>
<td>cine</td>
<td>['si.ne]</td>
<td>['θi.ne]</td>
<td>‘cinema’</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>cinco</td>
<td>['siŋ.ko]</td>
<td>['θiŋ.ko]</td>
<td>‘five’</td>
</tr>
<tr>
<td>/i/</td>
<td></td>
<td>cintura</td>
<td>[sin.ˈtu.ɾa]</td>
<td>[θin.ˈtu.ɾa]</td>
<td>‘waist’</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>citado</td>
<td>[si.ˈta.ɾo]</td>
<td>[θi.ˈta.ɾo]</td>
<td>‘cited’</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>decir</td>
<td>[de.ˈʃir]</td>
<td>[de.ˈθir]</td>
<td>‘to say’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cocina</td>
<td>[ko.ˈsi.na]</td>
<td>[ko.ˈθi.na]</td>
<td>‘kitchen’</td>
</tr>
<tr>
<td>medial</td>
<td>yes</td>
<td>principal</td>
<td>[prin.ˈsi.ˈpal]</td>
<td>[prin.ˈθi.ˈpal]</td>
<td>‘main’</td>
</tr>
<tr>
<td></td>
<td>no</td>
<td>difícil</td>
<td>[di.ˈfi.ɾil]</td>
<td>[di.ˈθi.θil]</td>
<td>‘difficult’</td>
</tr>
</tbody>
</table>
### Dialectal Difference #2 (BAS /ʃ/ vs. MS /ʝ/): orthographic <y>

#### Table 25: Stimuli for Dialectal Difference #2: orthographic <y>

<table>
<thead>
<tr>
<th>vowel</th>
<th>position</th>
<th>stress</th>
<th>token</th>
<th>BAS</th>
<th>MS</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>initial</td>
<td>yes</td>
<td>yate</td>
<td>[ʃa.te], [ʒa.te]</td>
<td>[ʃa.te]</td>
<td>'yacht'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>yacimiento</td>
<td>[ʃa.si.'mjen.to], [ʒa.si.'mjen.to]</td>
<td>[ʃa.θi.'mjen.to]</td>
<td>'site'</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>apoyar</td>
<td>[a.po.'ʃar], [a.po.'ʒar]</td>
<td>[a.po.'ʃar]</td>
<td>'to support'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>raya</td>
<td>[ra.ʃa], [ra.ʒa]</td>
<td>[ra.ʃa]</td>
<td>'stripe'</td>
</tr>
<tr>
<td>/e/</td>
<td>initial</td>
<td>yes</td>
<td>yendo</td>
<td>[ʃen.do], [ʒen.do]</td>
<td>[ʃen.do]</td>
<td>'going'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>yeismo</td>
<td>[ʃe.'iz.mo], [ʒe.'iz.mo]</td>
<td>[ʃe.'iz.mo]</td>
<td>'yeismo'</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>proyecto</td>
<td>[ʃro.'ʃek.to], [ʃro.'ʒek.to]</td>
<td>[ʃro.'ʃek.to]</td>
<td>'project'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>bueyes</td>
<td>[ʃwe.ʃes], [ʃwe.ʒes]</td>
<td>[ʃwe.ʃes]</td>
<td>'oxen'</td>
</tr>
<tr>
<td>/i/</td>
<td>initial</td>
<td>yes</td>
<td>ensayista</td>
<td>[ʃen.sa.'ʃis.ta], [ʃen.sa.'ʒis.ta]</td>
<td>[ʃen.sa.'ʃis.ta]</td>
<td>'essayist'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/o/</td>
<td>initial</td>
<td>yes</td>
<td>yoga</td>
<td>[ʃo.ʃa], [ʒo.ʃa]</td>
<td>[ʃo.ʃa]</td>
<td>'yoga'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>yogur</td>
<td>[ʃo.'ʃur], [ʒo.'ʃur]</td>
<td>[ʃo.'ʃur]</td>
<td>'yogurt'</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>cayó</td>
<td>[ʃa.'ʃo], [ʒa.'ʃo]</td>
<td>[ʃa.'ʃo]</td>
<td>'fell'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>mayoría</td>
<td>[ʃma.ʃo.'ri.a], [ʃma.ʒo.'ri.a]</td>
<td>[ʃma.ʃo.'ri.a]</td>
<td>'majority'</td>
</tr>
<tr>
<td>/u/</td>
<td>initial</td>
<td>yes</td>
<td>yunque</td>
<td>[ʃun.ʃe], [ʒun.ʃe]</td>
<td>[ʃun.ʃe]</td>
<td>'anvil'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>yugoslava</td>
<td>[ʃu.ʒos.'la.βa], [ʒu.ʒos.'la.βa]</td>
<td>[ʃu.ʒos.'la.βa]</td>
<td>'Yugoslavian'.fem</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>ayuda</td>
<td>[ʃa.'ʃu.ʃa], [ʒa.'ʃu.ʃa]</td>
<td>[ʃa.'ʃu.ʃa]</td>
<td>'help'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>ayudado</td>
<td>[ʃa.ʃu.'do.ʃa], [ʒa.ʃu.'do.ʃa]</td>
<td>[ʃa.ʃu.'do.ʃa]</td>
<td>'help'.pastpart</td>
</tr>
</tbody>
</table>
Dialectal Difference #2 (BAS /ʃ/ vs. MS /ʝ/): orthographic <ll>

<table>
<thead>
<tr>
<th>vowel</th>
<th>position</th>
<th>stress</th>
<th>token</th>
<th>BAS</th>
<th>MS</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>initial</td>
<td>yes</td>
<td>llamо</td>
<td>['ʃa.mo'], ['ʒa.mo']</td>
<td>['ja.mo']</td>
<td>'I call'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>llamаr</td>
<td>['ʃa.'mar'], ['ʒa.'mar']</td>
<td>['ja.'mar']</td>
<td>'to call'</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>callаd о</td>
<td>['ka.'ja.ðo'], ['ka.'ʒa.ðo']</td>
<td>['ka.'ja.ðo']</td>
<td>'quiet'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>villa</td>
<td>['bi.ʃa'], ['bi.ʒa']</td>
<td>['bi.ʝa']</td>
<td>'village'</td>
</tr>
<tr>
<td>/e/</td>
<td>initial</td>
<td>yes</td>
<td>llegа</td>
<td>['ʃe.ɣa'], ['ʒe.ɣa']</td>
<td>['ʝe.ɣa']</td>
<td>'arrive'.3sg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>llevar</td>
<td>['ʃe.'βar'], ['ʒe.'βar']</td>
<td>['ʝe.'βar']</td>
<td>'to carry'</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>gallegо</td>
<td>['ga.'ʃe.yo'], ['ga.'ʒe.yo']</td>
<td>['ga.'ʃe.yo']</td>
<td>'Galician'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>bulle</td>
<td>['bu.ʃe'], ['bu.ʒe']</td>
<td>['bu.ʝe']</td>
<td>'boil'.3sg</td>
</tr>
<tr>
<td>/i/</td>
<td>initial</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>gallina</td>
<td>[ga.'ʃi.na], [ga.'ʒi.na']</td>
<td>[ga.'ʃi.na']</td>
<td>'hen'</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/o/</td>
<td>initial</td>
<td>yes</td>
<td>llоро</td>
<td>['ʃo.ro'], ['ʒo.ro']</td>
<td>['jo.ro']</td>
<td>'I cry'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>llовия</td>
<td>['ʃo.'βi.a], ['ʒo.'βi.a']</td>
<td>['jo.'βi.a']</td>
<td>'it was raining'</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>millонеs</td>
<td>[mi.'ʃo.nes], [mi.'ʒo.nes']</td>
<td>[mi.'ʃo.nes']</td>
<td>'millions'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>cabалlo</td>
<td>[ka.'ʃa.ʃo], [ka.'ʒa.ʒo]</td>
<td>[ka.'ʃa.ʒo]</td>
<td>'horse'</td>
</tr>
<tr>
<td>/u/</td>
<td>initial</td>
<td>yes</td>
<td>llуvia</td>
<td>['ʃu.ʃja], ['ʒu.ʃja']</td>
<td>['ju.ʃja']</td>
<td>'rain'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>lluvidоs</td>
<td>[ʃu.'ʃjo.so], [ʒu.'ʃjo.so]</td>
<td>[ʃu.'ʃjo.so]</td>
<td>'rainy'</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>vellуdо</td>
<td>[be.'ʃu.ðo], [be.'ʒu.ðo]</td>
<td>[be.'ʃu.ðo]</td>
<td>'hairy'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dialectal Difference #3 (BAS Laminal [s] vs. MS Apical [ʃ]): orthographic <s>

Table 27: Stimuli for Dialectal Difference #3

<table>
<thead>
<tr>
<th>Vowel position</th>
<th>Vowel</th>
<th>stress</th>
<th>token</th>
<th>BAS</th>
<th>MS</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/ initial</td>
<td>/a/</td>
<td>yes</td>
<td>sala</td>
<td>[ˈsa.la]</td>
<td>[ˈʃa.la]</td>
<td>‘classroom’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>salir</td>
<td>[ˈsa.l̞iɾ]</td>
<td>[ˈʃa.l̞iɾ]</td>
<td>‘to leave’</td>
</tr>
<tr>
<td>medial</td>
<td></td>
<td>yes</td>
<td>pasado</td>
<td>[pə.'sa.ðo]</td>
<td>[pə.'ʃa.ðo]</td>
<td>‘past’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>casa</td>
<td>[ˈka.s̜a]</td>
<td>[ˈka.ʃa]</td>
<td>‘house’</td>
</tr>
<tr>
<td>/e/ initial</td>
<td>/e/</td>
<td>yes</td>
<td>seda</td>
<td>[ˈse.ða]</td>
<td>[ˈʃe.ða]</td>
<td>‘silk’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>semanas</td>
<td>[se.'ma.nas]</td>
<td>[ʃe.'ma.nas]</td>
<td>‘weeks’</td>
</tr>
<tr>
<td>medial</td>
<td></td>
<td>yes</td>
<td>diseño</td>
<td>[di.'se.ɲo]</td>
<td>[di.'ʃe.ɲo]</td>
<td>‘design’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>clase</td>
<td>[ˈ kla.s̜e]</td>
<td>[ˈ kla.ʃe]</td>
<td>‘class’</td>
</tr>
<tr>
<td>/i/ initial</td>
<td>/i/</td>
<td>yes</td>
<td>silla</td>
<td>[ˈsi.f̊a]</td>
<td>[ˈʃi.f̊a]</td>
<td>‘chair’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>situado</td>
<td>[si.'tua.ðo]</td>
<td>[ʃi.'tua.ðo]</td>
<td>‘located’</td>
</tr>
<tr>
<td>medial</td>
<td></td>
<td>yes</td>
<td>posible</td>
<td>[po.'si.ʃle]</td>
<td>[po.'ʃi.ʃle]</td>
<td>‘possible’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>casi</td>
<td>[ˈka.si]</td>
<td>[ˈka.ʃi]</td>
<td>‘almost’</td>
</tr>
<tr>
<td>/o/ initial</td>
<td>/o/</td>
<td>yes</td>
<td>solo</td>
<td>[ˈso.lo]</td>
<td>[ˈʃo.lo]</td>
<td>‘alone’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>sonido</td>
<td>[so.'ni.ðo]</td>
<td>[ʃo.'ni.ðo]</td>
<td>‘sound’</td>
</tr>
<tr>
<td>medial</td>
<td></td>
<td>yes</td>
<td>persona</td>
<td>[per.'so.na]</td>
<td>[per.'ʃo.na]</td>
<td>‘person’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>caso</td>
<td>[ˈka.so]</td>
<td>[ˈka.ʃo]</td>
<td>‘case’</td>
</tr>
<tr>
<td>/u/ initial</td>
<td>/u/</td>
<td>yes</td>
<td>suma</td>
<td>[ˈsu.ma]</td>
<td>[ˈʃu.ma]</td>
<td>‘sum’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>sujeto</td>
<td>[su.'xe.to]</td>
<td>[ʃu.'xe.to]</td>
<td>‘subject’</td>
</tr>
<tr>
<td>medial</td>
<td></td>
<td>yes</td>
<td>consumo</td>
<td>[kon.'su.mo]</td>
<td>[kon.'ʃu.mo]</td>
<td>‘consumption’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no</td>
<td>resultó</td>
<td>[re.sul.'to]</td>
<td>[re.ʃul.'to]</td>
<td>‘it turned out’</td>
</tr>
</tbody>
</table>
Table 28: Stimuli for Dialectal Difference #4

<table>
<thead>
<tr>
<th>Vowel</th>
<th>position</th>
<th>stress</th>
<th>token</th>
<th>BAS</th>
<th>MS</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>initial</td>
<td>yes</td>
<td>jarrə</td>
<td>['xa.ra]</td>
<td>['χa.ra]</td>
<td>‘classroom’</td>
</tr>
<tr>
<td></td>
<td>initial</td>
<td>no</td>
<td>jamás</td>
<td>[xa.'mas]</td>
<td>[χa.'mas]</td>
<td>‘to leave’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>dejado</td>
<td>[de.'xa.do]</td>
<td>[de.'χa.do]</td>
<td>‘past’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>no</td>
<td>hija</td>
<td>['i.xa]</td>
<td>['i.χa]</td>
<td>‘house’</td>
</tr>
<tr>
<td>/e/</td>
<td>initial</td>
<td>yes</td>
<td>jefe</td>
<td>[xe.fe]</td>
<td>[χe.fe]</td>
<td>‘silk’</td>
</tr>
<tr>
<td></td>
<td>initial</td>
<td>no</td>
<td>Jesus</td>
<td>[xe.'sus]</td>
<td>[χe.'sus]</td>
<td>‘weeks’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>mujer</td>
<td>[mu.'xer]</td>
<td>[mu.'χer]</td>
<td>‘design’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>no</td>
<td>viaje</td>
<td>['bja.xe]</td>
<td>['bja.χe]</td>
<td>‘class’</td>
</tr>
<tr>
<td>/i/</td>
<td>initial</td>
<td>yes</td>
<td>gira</td>
<td>[xi.ra]</td>
<td>[χi.ra]</td>
<td>‘chair’</td>
</tr>
<tr>
<td></td>
<td>initial</td>
<td>no</td>
<td>gigante</td>
<td>[xi.'yan.te]</td>
<td>[χi.'yan.te]</td>
<td>‘located’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>mejilla</td>
<td>[me.'xi.ja]</td>
<td>[me.'χi.ja]</td>
<td>‘possible’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>no</td>
<td>mejicano</td>
<td>[me.xi.'ka.no]</td>
<td>[me.xi.'ka.no]</td>
<td>‘almost’</td>
</tr>
<tr>
<td>/o/</td>
<td>initial</td>
<td>yes</td>
<td>joven</td>
<td>[xo.'ben]</td>
<td>[χo.'ben]</td>
<td>‘alone’</td>
</tr>
<tr>
<td></td>
<td>initial</td>
<td>no</td>
<td>José</td>
<td>[xo.'se]</td>
<td>[χo.'se]</td>
<td>‘sound’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>mejor</td>
<td>[me.'xor]</td>
<td>[me.'χor]</td>
<td>‘person’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>no</td>
<td>ajo</td>
<td>['a.xo]</td>
<td>['a.χo]</td>
<td>‘case’</td>
</tr>
<tr>
<td>/u/</td>
<td>initial</td>
<td>yes</td>
<td>julio</td>
<td>[xu.ljo]</td>
<td>[χu.ljo]</td>
<td>‘sum’</td>
</tr>
<tr>
<td></td>
<td>initial</td>
<td>no</td>
<td>jugar</td>
<td>[xu.'yar]</td>
<td>[χu.'yar]</td>
<td>‘subject’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>yes</td>
<td>injuria</td>
<td>[iŋ.'xu.rja]</td>
<td>[iŋ.'χu.rja]</td>
<td>‘consumption’</td>
</tr>
<tr>
<td></td>
<td>medial</td>
<td>no</td>
<td>ajustar</td>
<td>[a.xus.'tar]</td>
<td>[a.χus.'tar]</td>
<td>‘it turned out’</td>
</tr>
</tbody>
</table>
Table 29: *Stimuli for Dialectal Difference #5*

<table>
<thead>
<tr>
<th>sequence</th>
<th>position</th>
<th>stress</th>
<th>token</th>
<th>BAS</th>
<th>MS</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>/io/</td>
<td>initial</td>
<td>pretonic</td>
<td>violin</td>
<td>[bjo.'lin]</td>
<td>[bjo.'lin]</td>
<td>'violin'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>violencia</td>
<td>[bjo.'len.sja]</td>
<td>[bjo.'len.θja]</td>
<td>'violence'</td>
</tr>
<tr>
<td></td>
<td>tonic</td>
<td>pretonic</td>
<td>nacional</td>
<td>[djo.'ses]</td>
<td>[djo.θeg]</td>
<td>'gods'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>curiosidad</td>
<td>[ra.sjo.θal]</td>
<td>[ra.θo.nal]</td>
<td>'national'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mansiones</td>
<td>[man.'sjoe.s]</td>
<td>[man.θjo.neg]</td>
<td>'mansions'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>misiones</td>
<td>[mi.'sjoe.s]</td>
<td>[mi.θjo.neg]</td>
<td>'missions'</td>
</tr>
<tr>
<td></td>
<td>posttonic</td>
<td></td>
<td>medio</td>
<td>['me.ðjo]</td>
<td>['me.ðjo]</td>
<td>'half'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>principio</td>
<td>[prin.'si.pjo]</td>
<td>[prin.'θi.pjo]</td>
<td>'beginning'</td>
</tr>
<tr>
<td>/eo/</td>
<td>initial</td>
<td>pretonic</td>
<td>leopardo</td>
<td>[ljo.'pa.θo]</td>
<td>[le.o.'pa.θo]</td>
<td>'leopard'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>teoría</td>
<td>[tjo.'ri.a]</td>
<td>[te.o.'ri.a]</td>
<td>'theory'</td>
</tr>
<tr>
<td></td>
<td>tonic</td>
<td>pretonic</td>
<td>peones</td>
<td>[pjo.nes]</td>
<td>[p'e.o.nes]</td>
<td>'pawns'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>reojo</td>
<td>['rjo.xo]</td>
<td>['re.o.θo]</td>
<td>'corner of eye'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ideología</td>
<td>[i.'ðjo.lo.θi.a]</td>
<td>[i.'θe.o.lo.θi.a]</td>
<td>'ideology'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>campeonato</td>
<td>[kam.pjo.'na.to]</td>
<td>[kam.pe.o.'na.to]</td>
<td>'championship'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>acordeones</td>
<td>[a.kor.'ðjo.nes]</td>
<td>[a.kor.'θe.o.nes]</td>
<td>'accordions'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>antojos</td>
<td>[an.'tjo.xos]</td>
<td>[an.'te.o.θos]</td>
<td>'telescopes'</td>
</tr>
<tr>
<td></td>
<td>posttonic</td>
<td></td>
<td>petróleo</td>
<td>[pe.'tjo.le.o]</td>
<td>[pe.'tjo.θe.o]</td>
<td>'petroleum'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>cráneo</td>
<td>['kra.njo]</td>
<td>['kra.ne.o]</td>
<td>'skull'</td>
</tr>
<tr>
<td>/ia/</td>
<td>initial</td>
<td>pretonic</td>
<td>diamante</td>
<td>[dja.'man.te]</td>
<td>[dja.'man.te]</td>
<td>'diamante'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>criatura</td>
<td>[kra.'tu.ra]</td>
<td>[kra.'tu.ra]</td>
<td>'creature'</td>
</tr>
<tr>
<td></td>
<td>tonic</td>
<td></td>
<td>diario</td>
<td>['dja.rjo]</td>
<td>['dja.rjo]</td>
<td>'daily'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>viaje</td>
<td>['bja.xe]</td>
<td>['bja.xe]</td>
<td>'trip'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>aviador</td>
<td>[a.'βja.θor]</td>
<td>[a.'βja.θor]</td>
<td>'pilot'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>especialmente</td>
<td>[es.pe.sjal.'men.te]</td>
<td>[es.pe.'θjal.'men.te]</td>
<td>'especially'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Indiana</td>
<td>[in.'dja.na]</td>
<td>[in.'dja.na]</td>
<td>'Indiana'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>social</td>
<td>[so.'sjal]</td>
<td>[so.'θal]</td>
<td>'social'</td>
</tr>
<tr>
<td></td>
<td>posttonic</td>
<td></td>
<td>copia</td>
<td>['ko.pja]</td>
<td>['ko.pja]</td>
<td>'copy'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>hacia</td>
<td>['a.sja]</td>
<td>['a.θja]</td>
<td>'towards'</td>
</tr>
<tr>
<td>/ea/</td>
<td>initial</td>
<td>pretonic</td>
<td>peatonal</td>
<td>[pja.to.'nal]</td>
<td>[pe.a.to.'nal]</td>
<td>'pedestrian'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>realidad</td>
<td>[rja.lj.'θaθ]</td>
<td>[re.a.li.'θaθ]</td>
<td>'reality'</td>
</tr>
<tr>
<td></td>
<td>tonic</td>
<td></td>
<td>teatro</td>
<td>['tja.tro]</td>
<td>['te.a.tro]</td>
<td>'theatre'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>reales</td>
<td>['tja.les]</td>
<td>['re.a.leg]</td>
<td>'real'.pl</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>goleador</td>
<td>[gola.'θor]</td>
<td>[gola.e.a.'θor]</td>
<td>'scorer'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>planteamiento</td>
<td>[plan.tja.'mjen.to]</td>
<td>[plan.te.a.'mjen.to]</td>
<td>'approach'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lineal</td>
<td>[li.'njal]</td>
<td>[li.'ne.al]</td>
<td>'linear'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pelear</td>
<td>[pe.'ljar]</td>
<td>[pe.'le.ar]</td>
<td>'to fight'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>área</td>
<td>['a.rja]</td>
<td>['a.re.a]</td>
<td>'area'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linea</td>
<td>['li.nja]</td>
<td>['li.ne.a]</td>
<td>'line'</td>
</tr>
</tbody>
</table>
Dialectal Difference #6 (Potential Exceptional Hiatus in MS, but not in BAS)

Table 30: Stimuli for Dialectal Difference #6

<table>
<thead>
<tr>
<th>sequence</th>
<th>token</th>
<th>BAS</th>
<th>MS</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>duelo</td>
<td>['dwe.lo]</td>
<td>['dwe.lo]</td>
<td>'duel'</td>
</tr>
<tr>
<td></td>
<td>pie</td>
<td>['pje]</td>
<td>['pje]</td>
<td>'foot'</td>
</tr>
<tr>
<td></td>
<td>siendo</td>
<td>['sjen.do]</td>
<td>['sjen.do]</td>
<td>'being'</td>
</tr>
<tr>
<td></td>
<td>envidiando</td>
<td>[em.bi.'ðja.mos]</td>
<td>[em.bi.'ðja.mos]</td>
<td>'we envy'</td>
</tr>
<tr>
<td></td>
<td>fuimos</td>
<td>['fwi.mos]</td>
<td>['fwi.mos]</td>
<td>'we were/went'</td>
</tr>
<tr>
<td></td>
<td>diente</td>
<td>['djen.te]</td>
<td>['djen.te]</td>
<td>'tooth'</td>
</tr>
<tr>
<td></td>
<td>dio</td>
<td>['djo]</td>
<td>['djo]</td>
<td>'he/she gave'</td>
</tr>
<tr>
<td></td>
<td>cuida</td>
<td>['kwi.'ða]</td>
<td>['kwi.'ða]</td>
<td>'he/she looks after'</td>
</tr>
<tr>
<td></td>
<td>cuarto</td>
<td>['kwa.tɾo]</td>
<td>['kwa.tɾo]</td>
<td>'four'</td>
</tr>
<tr>
<td></td>
<td>cuando</td>
<td>['kwan.do]</td>
<td>['kwan.do]</td>
<td>'when'</td>
</tr>
<tr>
<td></td>
<td>viuda</td>
<td>['bju.'ða]</td>
<td>['bju.'ða]</td>
<td>'widow'</td>
</tr>
<tr>
<td></td>
<td>dueto</td>
<td>['dwe.to]</td>
<td>['du.'e.to]</td>
<td>'duet'</td>
</tr>
<tr>
<td></td>
<td>pió</td>
<td>['pi.'e]</td>
<td>['pi.'e]</td>
<td>'I chirped'</td>
</tr>
<tr>
<td></td>
<td>riendo</td>
<td>['rjen.do]</td>
<td>['ri.'en.do]</td>
<td>'laughing'</td>
</tr>
<tr>
<td></td>
<td>enviando</td>
<td>[em.'bja.mos]</td>
<td>[em.bi.'a.mos]</td>
<td>'we send'</td>
</tr>
<tr>
<td></td>
<td>huimos</td>
<td>['wi.mos]</td>
<td>['u.'i.mos]</td>
<td>'we flee'</td>
</tr>
<tr>
<td></td>
<td>riel</td>
<td>['ri.'el]</td>
<td>['ri.'el]</td>
<td>'rail'</td>
</tr>
<tr>
<td></td>
<td>cliente</td>
<td>['kli.'en.te]</td>
<td>['kli.'en.te]</td>
<td>'client'</td>
</tr>
<tr>
<td></td>
<td>rió</td>
<td>['ri.'o]</td>
<td>['ri.'o]</td>
<td>'he/she laughed'</td>
</tr>
<tr>
<td></td>
<td>huída</td>
<td>['wi.'ða]</td>
<td>['u.'i.'ða]</td>
<td>'escape, flight'</td>
</tr>
<tr>
<td></td>
<td>situado</td>
<td>[si.'twa.'ðo]</td>
<td>[si.tu.'a.'ðo]</td>
<td>'located'</td>
</tr>
<tr>
<td></td>
<td>Rwanda</td>
<td>['rwan.da]</td>
<td>[ru.'an.da]</td>
<td>'Rwanda'</td>
</tr>
<tr>
<td></td>
<td>diurno</td>
<td>['djur.no]</td>
<td>[di.'ur.no]</td>
<td>'daytime.adj'</td>
</tr>
</tbody>
</table>
Appendix B – Information Letter and Consent Form

Spanish

Departamento de Lingüística
Universidad de Toronto

Otoño de 2010

Estimado participante:

Soy estudiante de doctorado del departamento de lingüística. Estoy realizando una investigación sobre las variedades de español hablado en Buenos Aires, Argentina y en Madrid, España. Para completar mi trabajo, necesito la colaboración de hablantes nativos del español de estas regiones.

La participación en este estudio incluye pronunciar palabras, escuchar sonidos y participar en una conversación de aproximadamente 30 minutos con otro participante. La lectura va a ser grabada. Se le pedirá también que complete un cuestionario breve con algunos datos acerca de su conocimiento de otros idiomas u otros dialectos del español y de sus experiencias en Madrid.

Todos los materiales van a ser tratados en forma confidencial. Sólo la investigadora involucrada en este estudio y su supervisora van a tener acceso al material, el cual será codificado mediante la asignación de un número a cada participante. El cuestionario y los datos se conservarán por separado de cualquier otro formulario que contenga su nombre e información personal. Los nombres de los participantes no serán revelados en ninguna comunicación de este estudio.

Si tiene alguna pregunta sobre sus derechos como participante en este experimento, puede contactarme a mí, a mi supervisora o al Office of Research Ethics por correo electrónico al ethics.review@utoronto.ca o por teléfono al 416-946-3273.

Su participación en este estudio es voluntaria y puede interrumpirla en cualquier momento. El estudio durará aproximadamente dos horas y recibirá veinte euros en compensación por su colaboración.

Le agradezco su voluntad de participar. Para indicar su aceptación de participar y de ser grabada, por favor, firme el formulario adjunto. Si tiene alguna pregunta, no dude en hacérmela ahora o en comunicarse conmigo después de su participación.

Un cordial saludo,

Beth MacLeod,     Yoonjung Kang
Estudiante de Doctorado, Depto. de Lingüística  Profesora Agregada, Depto. de Lingüística
beth.macleod@utoronto.ca    kang@utsc.utoronto.ca
FORMULARIO DE ACEPTACIÓN

Yo, el abajo firmante, he sido informado de la naturaleza de este estudio, incluyendo la tarea que voy a hacer y accedo a participar. Entiendo que la lectura será grabada y accedo a esto. Entiendo que mi participación es voluntaria y puede ser interrumpida en cualquier momento.

Firma:

Aclaración (en imprenta):

Fecha:
Dear Participant:

I am a PhD student in the Department of Linguistics studying the variety of Spanish spoken in Buenos Aires, Argentina and in Madrid, Spain. Part of my research involves conducting an experimental study in which native speakers of Spanish from these areas are needed to participate.

Participation in this study will involve reading a paragraph and participating in a conversation of approximately 30 minutes in duration with another participant. The experiment will be recorded. You will also be asked to complete a questionnaire concerning your linguistic background and discuss your experiences in Madrid with the researcher.

Please note that all materials will be treated confidentially. My supervisor and I alone will have access to the forms and data, both of which will be coded with a participant number. The questionnaire and data will be stored separately from any form containing your name and personal information and will contain no links to such forms. Names of participants will never be revealed in any reports of this study.

If you have any questions about your rights as a participant in this study you may contact the Ethics Review Office at ethics.review@utoronto.ca or 416-946-3273.

Participation in this study is voluntary. You may withdraw at any time. Total time involved will be approximately 2 hours. You will receive €20 for your participation.

We very much appreciate your willingness to participate. To indicate your consent to participate in the study and to be recorded, please sign the attached form. If you have any questions, please do not hesitate to ask them now or to contact me subsequent to your participation.

Yours sincerely,

Beth MacLeod, PhD Student
Department of Linguistics, University of Toronto
beth.macleod@utoronto.ca

Yoonjung Kang, Assistant Professor
Department of Linguistics, University of Toronto
kang@utsc.utoronto.ca
CONSENT FORM

I, the undersigned, have been informed of the nature of the present study, including the tasks to be undertaken, and agree to participate. I agree to have myself recorded as I read the sentences in this study. I understand that my participation is voluntary and that I may withdraw at any time.

Name (please print):

Signature:

Date:
Appendix C – Background Questionnaire

Spanish – Buenos Aires Spanish speakers

1. Información Personal
   • Sexo: Masculino  Femenino
   • Año de Nacimiento:
   • Lugar de Nacimiento: Ciudad/Pueblo  País
   • ¿En qué ciudad/pueblo creciste?
   • ¿Has vivido en alguna otra ciudad de tu país o del mundo por más de un año?
     Sí  No
     En caso de una respuesta afirmativa, por favor, indica el lugar:
   • ¿En qué ciudad/pueblo creciste?
   • Ocupación:
   • Nivel de Educación Completado:  Secundaria  Universidad  Profesional
   • ¿Estás cursando alguna carrera universitaria?  Sí  No

2. Madrid y Buenos Aires
   • ¿Cuánto tiempo llevas en Madrid?
   • ¿Vives en Madrid o estás de visita?
     Si vives en Madrid, ¿por qué te mudaste aquí?
   • ¿Piensas quedarte en Madrid?
   • ¿Te gusta Madrid?
   • ¿Qué tal la gente madrileña?
   • ¿Se notan algunas diferencias entre la gente de Madrid y la de Buenos Aires?
   • ¿Tienes muchos amigos de Buenos Aires en Madrid?
   • ¿Te gusta Buenos Aires?

3. El Español
   • Hay muchas diferencias entre el español de Madrid y el de Buenos Aires. En términos de la
     pronunciación (y no de las palabras en sí), ¿cuáles diferencias has notado?
   • ¿Te gusta el acento español?  ¿Prefieres el acento español o argentino?
   • Muchas veces cuando alguien se muda a otro país, la manera de hablar le cambia. ¿Has notado
     que tu español ha cambiado desde que llegaste a Madrid?
   • ¿Quieres que tu español cambie o que no cambie?  ¿Por qué?
   • Si tuvieras hijos en Madrid, ¿te molestaría si hablan con el acento español?
1. Información Personal
   • Sexo: Masculino Femenino
   • Año de Nacimiento:
   • Lugar de Nacimiento: Ciudad/Pueblo País
   • ¿En qué ciudad/pueblo creciste?
   • ¿Has vivido en alguna otra ciudad de tu país o del mundo por más de un año? 
     Sí No
     En caso de una respuesta afirmativa, por favor, indica el lugar:
   • Ocupación:
   • Nivel de Educación Completado: Secundaria Universidad Profesional
   • Estas cursando alguna carrera universitaria? Sí No

2. Madrid y Buenos Aires
   • ¿Te gusta Madrid?
   • ¿Has estado en Buenos Aires?
   • ¿Conoces a gente de Buenos Aires (o de Argentina) aquí en Madrid?
   • ¿Crees que hay algunas diferencias entre la gente de Madrid y la gente de Argentina?
   • ¿Te parece que la gente de Argentina es bienvenida en Madrid? ¿Se nota alguna tensión entre las comunidades argentina y española?

3. El Español
   • Hay muchas diferencias entre el español de Madrid y el de Buenos Aires. En términos de la pronunciación (y no de las palabras en sí), ¿cuáles diferencias has notado?
   • ¿Te gusta el acento argentino? ¿Prefieres el acento español o el argentino?
   • Muchas veces cuando alguien se muda a otro país, la manera de hablar le cambia. ¿Si te mudaras a Argentina, querrías que tu español cambie o que no cambie? ¿Por qué?
1. Personal Information
   - Sex: Male  Female
   - Year of Birth:
   - Place of Birth: City/Town  Country
   - Which city/town did you grow up in?
   - Have you ever lived in another city or country for more than 1 year?
     Yes  No
     If yes, please state where:
   - Occupation:
   - Level of Education Completed: Secondary  University  Professional
   - Are you working on a university education currently? Yes  No

2. Madrid and Buenos Aires
   - How long have you been in Madrid?
   - Do you live in Madrid or are you visiting?
   - If you live in Madrid, why did you move here?
   - Are you planning to stay in Madrid?
   - Do you like Madrid?
   - How do you like the people in Madrid?
   - Have you noticed any differences between the people in Madrid and the people in Buenos Aires?
   - Do you have many friends from Buenos Aires in Madrid?
   - Do you like Buenos Aires?

3. Spanish
   - There are many differences between the Spanish of Buenos Aires and the Spanish of Madrid. In terms of the pronunciation (and not in terms of the words, as such), which differences have you noticed?
   - Do you like the accent of the Spanish in Madrid? Do you prefer the Madrid or the Buenos Aires accent?
   - Many times when someone moves to another country the way they speak changes. Have you noticed any changes in your Spanish since you arrived in Madrid?
   - Do you want your Spanish to change or not to change? Why?
   - If you had children in Madrid, would it bother you if they spoke with the Madrid accent?
1. Personal Information
   - Sex: Male Female
   - Year of Birth:
   - Place of Birth: City/Town Country
   - Which city/town did you grow up in?
   - Have you ever lived in another city or country for more than 1 year?
     Yes No
     If yes, please state where:
   - Occupation:
   - Level of Education Completed: Secondary University Professional
   - Are you working on a university education currently? Yes No

2. Madrid and Buenos Aires
   - Do you like Madrid?
   - Have you ever been to Buenos Aires?
   - Do you know people from Buenos Aires (or Argentina) here in Madrid?
   - Do you think there are any differences between the people of Madrid and the people of Argentina?
   - Do you think people from Argentina are welcome in Madrid? Do you notice any tension between the two communities?

3. Spanish
   - There are many differences between the Spanish of Buenos Aires and the Spanish of Madrid. In terms of the pronunciation (and not in terms of the words, as such), which differences have you noticed?
   - Do you like the accent of the Spanish in Madrid? Do you prefer the Madrid or the Buenos Aires accent?
   - Many times when someone moves to another country the way they speak changes. If you moved to Argentina, would you want your Spanish to stay the same or change? Why?
Appendix D – Statistical Analysis of Perception/Repetition Task Stimuli

Dialectal Difference #1 (BAS /s/ versus MS /θ/)  

Table 31: Model summary: linear mixed effects model with relative intensity as dependant.  
Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coef. Est.</th>
<th>StdErr</th>
<th>t-value</th>
<th>p-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-13.500</td>
<td>4.196</td>
<td>-3.217</td>
<td>0.01</td>
<td>*</td>
</tr>
<tr>
<td>DIALECT_MS</td>
<td>-9.833</td>
<td>3.885</td>
<td>-2.531</td>
<td>0.03</td>
<td>*</td>
</tr>
</tbody>
</table>

Dialectal Difference #2 (BAS /ʃ/ versus MS /ʝ/)  

Table 32: Model summary: linear mixed effects model with relative intensity as dependant.  
Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coef. Est.</th>
<th>StdErr</th>
<th>t-value</th>
<th>p-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-16.500</td>
<td>3.377</td>
<td>-4.886</td>
<td>&lt;0.0001</td>
<td>***</td>
</tr>
<tr>
<td>DIALECT_MS</td>
<td>7.071</td>
<td>3.396</td>
<td>2.082</td>
<td>0.05</td>
<td>*</td>
</tr>
</tbody>
</table>

Dialectal Difference #3 (BAS [s] versus MS [ʂ])  

Table 33: Model summary: linear mixed effects model with centre of gravity as dependant.  
Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coef. Est.</th>
<th>StdErr</th>
<th>t-value</th>
<th>p-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>7718.7</td>
<td>981.6</td>
<td>7.863</td>
<td>&lt;0.0001</td>
<td>***</td>
</tr>
<tr>
<td>DIALECT_MS</td>
<td>-2407.2</td>
<td>908.8</td>
<td>-2.649</td>
<td>0.02</td>
<td>*</td>
</tr>
</tbody>
</table>

Dialectal Difference #4 (BAS [x] versus MS [χ])  

Table 34: Model summary: linear mixed effects model with F2 at the onset of the following vowel as dependant.  
Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coef. Est.</th>
<th>StdErr</th>
<th>t-value</th>
<th>p-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3321.4</td>
<td>117.9</td>
<td>28.168</td>
<td>&lt;0.0001</td>
<td>***</td>
</tr>
<tr>
<td>DIALECT_MS</td>
<td>-216.4</td>
<td>109.2</td>
<td>-1.982</td>
<td>0.073</td>
<td>.</td>
</tr>
</tbody>
</table>
Dialectal Difference #5 (diphthongization of mid vowel sequences in BAS, not in MS)

Table 35: Model summary: linear mixed effects model with composite measure (sequence duration and F1) as dependent. Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coef. Est.</th>
<th>StdErr</th>
<th>t-value</th>
<th>p-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.3030</td>
<td>0.2294</td>
<td>1.321</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>DIALECT_MS</td>
<td>0.7465</td>
<td>0.2124</td>
<td>3.515</td>
<td>0.003</td>
<td>**</td>
</tr>
</tbody>
</table>

Dialectal Difference #6 (potential exceptional hiatus in MS, not in BAS)

Table 36: Model summary: linear mixed effects model with sequence duration as dependant. Significance codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coef. Est.</th>
<th>StdErr</th>
<th>t-value</th>
<th>p-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.17918</td>
<td>0.03195</td>
<td>5.608</td>
<td>&lt;0.0001</td>
<td>***</td>
</tr>
<tr>
<td>DIALECT_MS</td>
<td>0.06362</td>
<td>0.02958</td>
<td>2.151</td>
<td>0.048</td>
<td>*</td>
</tr>
</tbody>
</table>
## Appendix E – Raw Data for Pre and Post-Conversation Production and Perception

<table>
<thead>
<tr>
<th>Diff #1: /s/-/θ/ Relative Intensity</th>
<th>Diff #2: /ʃ/-/j/ Relative Intensity</th>
<th>Diff #3: /s/ Centre of Gravity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair</td>
<td>Dialect</td>
<td>pre mean</td>
</tr>
<tr>
<td>1</td>
<td>BAS</td>
<td>-20.2</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-21.1</td>
</tr>
<tr>
<td>2</td>
<td>BAS</td>
<td>-15.6</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-21.3</td>
</tr>
<tr>
<td>3</td>
<td>BAS</td>
<td>-12.2</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-20.7</td>
</tr>
<tr>
<td>4</td>
<td>BAS</td>
<td>-15.6</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-21.3</td>
</tr>
<tr>
<td>5</td>
<td>BAS</td>
<td>-12.2</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-20.7</td>
</tr>
<tr>
<td>6</td>
<td>BAS</td>
<td>-15.6</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-21.3</td>
</tr>
<tr>
<td>7</td>
<td>BAS</td>
<td>-17.9</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-25.2</td>
</tr>
<tr>
<td>8</td>
<td>BAS</td>
<td>-21.5</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-16.0</td>
</tr>
<tr>
<td>9</td>
<td>BAS</td>
<td>-16.0</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-23.1</td>
</tr>
<tr>
<td>10</td>
<td>BAS</td>
<td>-21.2</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-19.9</td>
</tr>
<tr>
<td>11</td>
<td>BAS</td>
<td>-7.8</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>-27.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diff #4: /x/</th>
<th>Diff #5: mids</th>
<th>Diff #6: hiatus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair</td>
<td>Dialect</td>
<td>F2Onset</td>
</tr>
<tr>
<td>1</td>
<td>BAS</td>
<td>2758</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>2973</td>
</tr>
<tr>
<td>2</td>
<td>BAS</td>
<td>2880</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>2583</td>
</tr>
<tr>
<td>3</td>
<td>BAS</td>
<td>2786</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>2610</td>
</tr>
<tr>
<td>4</td>
<td>BAS</td>
<td>2670</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>2586</td>
</tr>
<tr>
<td>5</td>
<td>BAS</td>
<td>2674</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>2698</td>
</tr>
<tr>
<td>6</td>
<td>BAS</td>
<td>2551</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>2739</td>
</tr>
<tr>
<td>7</td>
<td>BAS</td>
<td>2661</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>2691</td>
</tr>
<tr>
<td>8</td>
<td>BAS</td>
<td>2641</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>2699</td>
</tr>
<tr>
<td>9</td>
<td>BAS</td>
<td>2799</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>2637</td>
</tr>
<tr>
<td>10</td>
<td>BAS</td>
<td>3215</td>
</tr>
<tr>
<td></td>
<td>MS</td>
<td>2871</td>
</tr>
</tbody>
</table>

Relative Intensity Centre of Gravity